Perspective on ecological indicators

Daniel Goodman

Department of Ecology, Montana State University, Bozeman, MT 59717-0346, U.S.A.
E-mail: goodman@rapid.msu.montana.edu

The North Pacific Research Board issued a call for indicator science, and we are reminded that this will have to be solid science, because when it is used to make management decisions, the issues will be contentious. This is to say that the science needs to be solid enough to stand up in court.

The presentations these past days have painted a bewildering picture of the state of indicator science for the Bering Sea, because the Stock Assessment and Fishery Evaluation (SAFE) document reports 100+ indicators without a clear enough representation of which indicate what, how they are used, or how they could be used.

In fact, this list of 100+ indicators includes several different kinds of quantities, which complicates the message that they convey. What we have at the moment should be sorted so that we can categorize them in a way that clarifies what each is, what it does, and why we think it is worth measuring and reporting. This sidesteps, at least for the moment, the prescriptive definition of "what is an indicator" and instead, asks for a descriptive taxonomy of "what are the kinds of things that are in the present list of indictors in the SAFE document."

This way, rather than ask for a cosmic definition of a "good indicator," we can consider for each distinct kind of thing we are calling an indicator, what would constitute a good one within that class. This might serve as a basis for ranking priorities both for investment and for emphasis in communication. What might be grounds for dropping an indicator from the *Ecosystem Considerations* appendix of the SAFE report? What might be grounds for stopping monitoring one of those indicators?

The indicators reported in the SAFE report, as a set, are just time series of variables that have been measured. There seem to be three main kinds of measures:

1. direct measures of system state,
2. summaries of measures of system state,
3. surrogate measures of system state.

Examples of direct measures of system state might be chlorophyll, or ice, possibly measured at a defined set of locations, possibly reported as a spatial average. We think that these are informative in their own right.

Examples of summaries of measures of system state might be PDO or trophic level biomass ratios. Note that the summaries do not relieve the need for the underlying measurement. Mathematically, PDO is a linear combination of sea surface temperatures over a spatial field. We may believe scientifically that this is a very revealing way to describe climate state. But we still need to measure sea surface temperatures in order to calculate PDO. Fortunately there are lots of other reasons for measuring sea surface temperature, and it is now routinely done by means of remote sensors, so the marginal cost of obtaining a measurement of PDO is very small.

Surrogate measures are proxies for things that are too expensive to measure directly on a routine basis, but hopefully the proxy is well enough correlated with the object of our actual interest. For example, sea bird reproduction may correlate well with zooplankton production within a known radius of their rookery, and may be simpler to monitor than the zooplankton production itself.

The index of biotic integrity, used extensively in surveys for freshwater systems, is another example of a surrogate measure. The procedure for development and validation of this index is well documented. A set of reference sites ranging from degraded to pristine, within a defined habitat type and geographic region, are selected and sampled; easily identifiable biota are counted across that gradient; statistical operations identify a surrogate index, based on easy sampling of recognizable biota, that is correlated with the degree of degradation of the site. Once this calibration has been done, the index based on easy sampling can replace the possibly more difficult direct measurement of environmental stress at new sites that were not involved in the calibration. Different habitat types and geographic regions, of course, harbor different biota regardless, so for each habitat type and geographic region a distinct index of biotic integrity must be developed and calibrated.

What must we ask about these measures that we are calling indicators? The first question, as with any environmental measure is how well it is measured. There is often a serious amount of measurement error in the technology of measurement itself. Sampling error often is even more serious since, as a practical matter, the measurements may be made at a very limited number of times and places, yet the result may be taken to represent a field that is known to be very heterogeneous and which is known to exhibit large temporal variation.

When developing a surrogate index, another important issue arises. The surrogate is not credible as an index without documentation of the degree of correspondence with ground truth. A proposed index for which a ground truth is unmeasured or unmeasurable is not subject to validation. A surrogate for which a ground truth is not operationally defined as a measurable should be a non-starter. Note that ecosystem health is a metaphor, not a measurable.

The possible reasons for reporting an indicator are fourfold. The indicator may serve to quantify:

- utility,
- attainment,
- normalcy,
- forecast.

An example of an indicator that directly represents a measure of utility would be fisheries yield. Our interest in this is self-explanatory. The importance of an indicator with this motivation depends on the value of the utility that it measures.

An indicator of attainment is a measure of a quantity for which a management reference point has already been agreed upon. In one sense the importance of an indicator with this motivation is a social construct – it is as important as the agreement that is behind it. But if the agreement rested on a belief about ecological consequences (such as the amount of escapement necessary for maintenance of higher trophic levels, as in the international Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) agreement for managing fisheries in the Antarctic), the stability of the agreement may change with changes in scientific knowledge about the connection to consequences.

The interest in indicators of “normalcy” is based on the expectation that the system is unlikely to confront us with unwelcome surprises as long as the system is operating within known historic bounds. Thus the indicators of normalcy may be measures of state or rate or correlation for properties that we believe to be significant to system function, and for which we have a long enough historic record to have convincingly identified normal bounds.

The reasons for interest in a reliable forecast will depend on the quantity that is being predicted. It may be a description of system state, where the interest in the prediction is scientific. The quantity predicted may be of interest because it directly constitutes utility in its own right, as in forecasts of fishery yield, or the quantity being predicted may have broad ramifications, such as predictions of regime change or ecosystem upheaval.

Note that the claim for any of these reasons for interest in an indicator may merit a second look. If the claim is utility, is there wide acceptance that this is a measure of value? If the claim is attainment, is there an actual governing policy, and are the reasons for that policy sound? If the claim is normalcy, what are the defining boundaries for the normal envelope, and what is the empirical evidence for these boundaries? If the claim is forecast, what is the statistical confidence in the forecast, and what is the empirical basis underlying the calculation of confidence?

The preceding taxonomy of indicators suggests a descriptive definition of ecosystem indicators. An ecosystem indicator is something we can measure that in turn serves as a measure, an estimate, or a prediction of something we care about. In every case there is room to ask hard questions about how well we measure, how accurately we estimate, how reliably we predict, and why we care. It would be helpful if the 100+ indicators in the SAFE report were catalogued in this way, with an examination of the hard questions, and documentation of the available answers.

It may emerge that some of the hard questions cannot be answered very well for some subset of the indicators. In particular, it is imaginable that documentation for performance for some of the surrogates and predictors may be thin. If so, it is important to recognize that interest in these is speculative, so each must be treated as a scientific hypothesis which carries a scientific responsibility to test the hypothesis. Therefore, for any indicator which does not convincingly pass the first layer of hard questions, there should be a second layer of hard questions about how the hypothesis is being tested, what is the design of that test, and how ongoing measurements, monitoring both the indicator and the ground truth, will eventually resolve the hypothesis.