EMDS DECISION SUPPORT MODEL APPROACH
TO THE EVALUATION OF WATERSHED CONDITION

Prepared by
Sean Gordon & Kirsten Gallo
for the State-Federal Monitoring Partnership

Overview & Background

This paper discusses the potential utility of the Ecosystem Management Decision Support System (EMDS) decision support model approach in interagency coordination of monitoring and evaluation of watershed conditions. It is an evolving working draft being compiled mainly as answers to questions from the state-federal group as posed/interpreted by Steve Lanigan (USDA Forest Service). It is also designed to provide a basic conceptual introduction to this approach.

Partnership Objectives

- Help reconcile methods to evaluate aquatic conditions between federal and state agencies in the Northwest Forest Plan region
- Assess watershed health across a large landscape that includes state, private, and federal ownership

Rationale

- Provide consistency between different agency evaluations, especially in adjacent or overlapping areas
- Support dialog on different agency objectives and best methods for achieving them
- Possibly reduce survey time/effort, overlapping areas wouldn't need to be surveyed multiple times

Summary

The Ecosystem Management Decision Support System (EMDS) provides a flexible framework for constructing operational definitions of watershed condition derived directly from monitoring data. The main challenge with producing consistent watershed evaluations across agencies will likely be that differing institutional mandates and needs lead to different operational definitions. Using a common framework like EMDS could help agencies understand each other's definitions and serve as an aid to discussion, but it cannot itself harmonize these different objectives. Developing a "core" shared model is a worthy goal, but it is likely that exploring ways to coherently present multiple perspectives will also be needed. The basic trade-off between using one model to achieve broader non-technical understanding (public, courts, etc.) versus using multiple models to achieve greater technical accuracy for different agency objectives should be kept in mind.

Introduction

There are at least 2 different parts to any monitoring program:

1. data collection
2. data evaluation

The decision support model (DSM) used by the AREMP team is designed to assist primarily with the evaluation task. It consists of methods for comparing each piece of data to a reference condition and for synthesizing interpretations of different types of data (e.g. apples & oranges; large wood & pools).

The EMDS software provides a framework for building these models and linking them to georeferenced data, but it does not provide the models or thresholds. An analogy is the way that Excel provides a general spreadsheet framework and calculation functions but it's up to the author to provide the specific table structures and data.

EMDS models can also help calculate and prioritize the value of missing information. So, to the extent that the model captures all the monitoring objectives, it can also be used to drive the data collection task.

Within this DSM evaluation approach, there are two distinct levels of operations:

1. attribute level: compare each data point to a reference condition
2. model level: combine different attribute evaluations into an overall (synthetic) score

Since the approach tends to be goal-driven, it makes more sense to start with the combined result (model) and work backwards to the original data (attributes).

Model Level Q&A

The model level concerns how evaluations of different attributes are integrated, after each attribute has been evaluated. This is typically done by normalizing each attribute's evaluated score to a common range (e.g. -1 to 1), possibly weighting the different attributes (if some are known to be more important than others), and deciding how to combine them (can one compensate for another or is each one a limiting factor?).

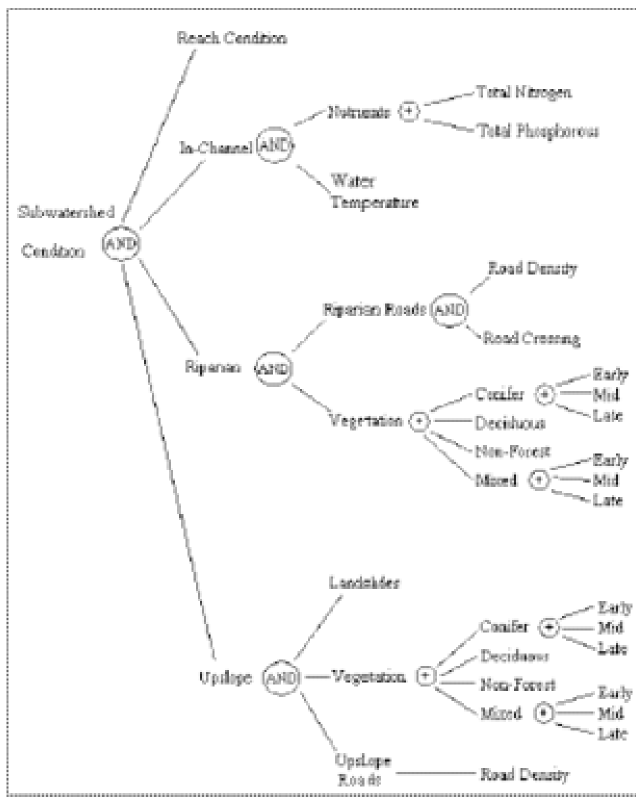

Figure 1. AREMP model structure

<u>Indicator Aggregation</u>

Why combine different attributes – why not just list each one separately?

Advantages of aggregation

- Most of the things we care about (watershed condition, water quality) consist of multiple attributes
- We often want more holistic assessments of these things
- Often there are too many attributes to make sense of if looked at individually

Disadvantages of aggregation

- If high scores in some attributes are allowed to compensate for low scores in others, then the combined result may mask these low scores. However, EMDS models do include specifications of how to control attribute combinations, including whether to take the average score ("+"), lowest ("AND") or highest ("OR")

Common Model

Should we develop one model with state/federal agencies that everyone agrees to use, or should each agency develop their own model? What are advantages/disadvantages to each approach? (Obtaining different results from the same data set because we are using different models would be a bad thing).

- As is roughly the case now, agencies use different models to characterize "watershed condition" (or related concepts). This leads to demands to explain why evaluations and subsequent decisions differ, and which model is most appropriate for the use (e.g. consultation and lawsuits). We may also have to prove that we did not choose the model that would give us "desired" results.
- Common models are useful and feasible to the extent that shared objectives exist. One objective cited by the group is the assessment of watershed health across a large landscape that includes state, private, and federal ownership.
- The main problem with trying to develop a shared model is that different agencies have different objectives, which in turn lead to different definitions of watershed health (one may focus on water quality for endangered fish while another is more focused on drinking water quality).
- Having different models producing different results is not necessarily a bad thing. To the extent the agency objectives are different, their model results should be too. The trade-off decision to be made appears to be between using one model to achieve broader non-technical understanding (public, courts, etc.) versus using multiple models to achieve greater technical accuracy for different agency objectives.
- Putting the models in a common framework facilitates an understanding of these differences.
- If agreeing on one overall watershed condition model is not feasible/practical, then agreement might still be possible on submodels (e.g. the physical water chemistry component)

Core Model

What about a "core model" that uses a single agreed upon subset of variables, but allows each agency to add on additional components? I suspect the state attributes will include attributes related to urban areas and estuaries that are irrelevant to the federal lands AREMP is analyzing. A core model again raises the issue of using different fuzzy curves if the protocols are different.
A core model with additional features which get turned on or off depending on the agency or objectives is quite possible using the EMDS framework. AREMP is using this approach to modify its model by physiographic province.

Missing Data

Assuming one model is developed, would the model "penalize" someone for not using all the attributes? E.g., State agencies don't collect periphyton data. Will they get "penalized" for not entering periphyton data? Also, I don't want to get in a situation where the states are collecting info on something related to urban areas for use in the model - but AREMP doesn't, e.g., impermeable surfaces - if AREMP is penalized in the model for a lack of data. Can the model be written so it doesn't penalize for some variables, but does for others? Related to the questions about "penalizing" for a lack of data, what does this really mean? My understanding is that even if

your watershed condition is ideal, you won't get a +1 condition rating is you are missing any data. Is this correct? Can the model be written so it is neutral about missing data?
Elements in models can easily be created so that they can be turned on or off, depending on the user's preference. So if periphyton is not to be used in the core model, it can be turned off and its score will not affect the total one way or the other. It can then be turned back on to run a more the complex (non-core) model.

Combining Data from Different Sampling Schemes

Assuming the same protocols are being used, what is feasibility of combining survey results from different sampling schemes using a DSM (e.g. EMAP uses only one sample point per watershed while the AREMP survey collects multiple points)?
Since the model expects only one value for each watershed-level attribute (landcover, roads, etc), the values from multiple surveys or samples would need to be averaged (significant discrepancies should probably be examined). Another solution is to use multiple scales of evaluation. For example, multiple sample points could be evaluated in a point layer and then the info summarized by watershed. In this case, all that is different is the metric, and a switch could be used to specify alternative methods for evaluating the info. Similarly, multiple values for an attribute from the same reach would need to be aggregated, but values from different reaches could be left as they are.

Evaluating Trends

How to evaluate results over time?

- Track trend in composite scores over time as a frequency distribution (can use straightforward statistical tests for evaluating changes in frequency distributions)
- Create a separate model to evaluate changes (Reynolds has used this approach for a Montreal Process indicators evaluation model, but generally recommends the first approach)

Attribute Level Q&A
At each node within the EMDS model, an attribute or a combination of attributes is evaluated. To begin with, data measurements are read in (water temperature, large wood, etc.). Each is then compared to an evaluation "fuzzy" curve and given a score between –1 and 1. This –1/1 score represents the degree to which the datum supports the overall proposition of the model. In this case, the overall model proposition is that the watershed is in good condition, and -1 means the data contradict this proposition while +1 would indicate complete support (0 would be neutral or no data). The evaluation curves are called "fuzzy" because they evaluate things using a continuum, instead of using a strict true/false cutoff threshold.

For example, let's say that one of the nodes involves evaluating the percent coverage of old growth forest (maybe as part of the Physical / Upslope / Vegetation node in Figure 1). The evaluation that would take place in this node could look like Figure 2.

The lighter curved line and left axis represent the probability distribution of the historical coverage of old growth forests in Oregon. Our best estimate is that variations in this coverage centered around 50%, although it may have been anywhere from 15% to 90%. Our current policy is likely to return coverage to around 25%, so socially we appear to have decided that this is fully sustainable (the risks and benefits of deviation from the reference condition are optimal). In this case, the evaluation curve in our decision support model might resemble the heavy line with its measurement scale on the right axis. If cover is between about 25% - 75%, we would evaluate it as fully sustainable (1.0), whereas less (or more) cover would receive declining scores. The reference condition could also be a habitat suitability index with the sustainability score reflecting a range of acceptable risk around the index.

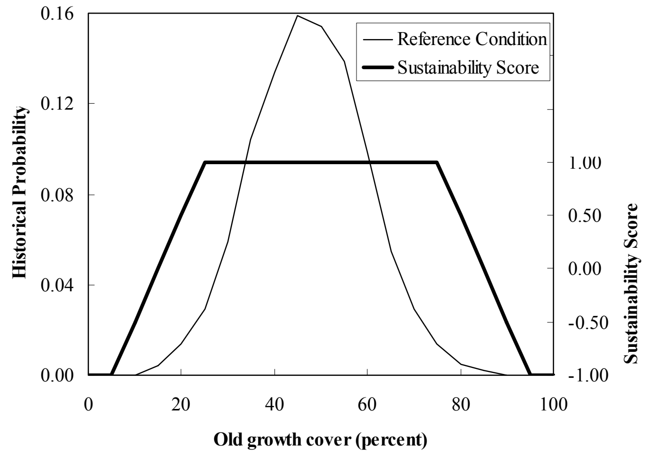

Figure 2. Data evaluation node

Reference Conditions

(Note: We have run into conflicting interpretations of the meaning of reference conditions and how they relate to other concepts such as sustainability targets, thresholds, etc. Hopefully the interagency process will bring more common understanding here, but until then be aware that the assumptions here may differ from your own.)

Why compare data to a reference condition?

Without some context or scale to compare it to, a lone data point is meaningless. Also, evaluation using a reference condition allows one to get the measurements for different attributes (water temperature, large wood) into a common scale (-1 to 1), so they can be combined.

How are reference conditions determined?

In this usage, a reference condition is not necessarily derived from a reference site
AREMP considered 3 common alternatives:

1. comparison to a reference site (one which exemplifies the condition sought)
2. comparison to a target (determined from ecological and/or social values)
3. comparison to past performance (has it gone up or down since the last measurement)

What is the effect of using different fuzzy curves for a given attribute? Is it legitimate to say the resulting watershed conditions are comparable?

In general, using different fuzzy curves to evaluate the same data will produce different results. The group is likely to start out in the position of each agency having both a different threshold (fuzzy curve) as well as a different data collection protocol for each attribute. This may not be as far apart as it sounds, if the objectives are similar. For example, one agency might measure average weekly temperature and have a corresponding threshold. Another might measure average weekly maximum temperatures but will probably have a threshold that is somewhat higher. If both sets of thresholds were designed to evaluate salmon habitat, results of these systems could be quite similar. In this case, conditions could be said to be comparable. However, since watershed condition tends to involve multiple attributes, ways of combining them, and data protocols, it would be difficult to make the case that two different methods are consistently comparable.

How in the world would one keep track of numerous fuzzy curves (and refining them over time)? The initial steps for this should not be very difficult technically. A spreadsheet or database that contains the basics (name of attribute, organization, thresholds, threshold sources, and associated datasets and sources) could be setup. Organizationally, it might be difficult to get updates and track them over time.

Data Collection

Are there any other benefits associated with using the same protocols for a given attribute?

- Consistency, especially in adjacent or overlapping areas
- Survey time, overlapping areas wouldn't need to be surveyed multiple times
- Data sharing would be facilitated

General Conclusions

The *process* of modeling can often be as important as the *product.* The process encourages:

- Problem definition: modeling forces thought about the objectives of monitoring; why are we collecting the info, how it will be used
- Accounting: modeling can provide a framework for how data are evaluated over time
- Sensitivity analysis: modeling can help determine the sensitivity of the overall result to changes in individual components (attributes)

Possible Next Steps

<u>Compiling agency evaluation criteria</u>

- Recommend adding separate columns/rows to the comparison documents which focus on what evaluation criteria are used (e.g. how does OR DEQ determine a stream is 303d?). Could put model level evaluations as a column in the Sampling Design document and attribute level criteria as rows in the Master comparison of protocols)

<u>Constructing test models</u>

- There are always devils in the implementation details, so it would help to work through a few example cases
- The best example would be the attribute for which the most agencies have evaluation criteria.
- The evaluation criteria for some indicators are more complex than others; starting with a simpler indicator would be best (e.g. wood in stream measures used in the AREMP model involve some of the more complex calculations, so a more straightforward attribute like temperature might make a better starting example).

Other Potentially Relevant Programs

- The Oregon Benchmarks program uses an indicator/target framework http://www.econ.state.or.us/opb/2001report/2001new.html
- Dutch Pressure-State-Response framework also a useful to think about how indicator variables differ

Related Partnership Resource Documents

Master comparison of agency protocols [SteveL 8/26/02]: Parameters X Agencies

Sample design comparisons [SteveL 8/26/02] : Objective, Attributes (see spreadsheet), Sample Design, Product/Report

BPA proposal for core indicators [Roy Beaty, BPA 8/27/02]

AREMP Monitoring Plan & Pilot Results [http://www.reo.gov/monitoring/reports.htm#watershed] see Knowledge-Based Decision-Support Model section

Other References

Reynolds, K. 2001. Prioritizing salmon habitat restoration with the AHP, SMART, and uncertain data. In: Schmoldt. The analytic hierarchy process in natural resource and environmental decision making. 199-217.

Reynolds, K.M. 2001. Fuzzy logic knowledge bases in integrated landscape assessment: examples and possibilities. General Technical Report PNW-GTR-521. Portland, OR : USDA Forest Service, Pacific Northwest Research Station.

Reynolds, K. and Peets, S. 2001. Integrated Assessment and Priorities for Protection and Restoration of Watersheds. Paper Presented at Forest Biometry, Modelling and Information Science, University of Greenwich, UK.

Reynolds, K. and Reeves, G.H. In press. Role of knowledge-based systems in analysis and communication of monitoring data. In: Busch, D.E. and Trexler, J.C. Interdisciplinary approaches to ecological monitoring of ecosystem initiatives. New York : Island Press.

Schultz, M.T. 2001. A critique of EPA's index of watershed indicators. Journal of Environmental Management 62(4): 429-443.

Contact

Sean Gordon
Ph.D. student, Forest Resources, OSU
503-582-9651 home/office
sean.gordon@orst.edu

Kirsten Gallo
Aquatic Ecologist
Aquatic and Riparian Effectiveness Monitoring Program
USDA Forest Service
541.750.7021
kgallo@fs.fed.us