Review of the TFW Monitoring Program:
Watershed-Scale Monitoring Pilot Project (Draft)
(Schuett-Hames, 1999)

Prepared by the Center for Streamside Studies
Thomas Sibley and Susan Bolton, editors
Additional contributors include, in alphabetical order,
Loveday Conquest, Rick Edwards, Martin Fox, Nancy Gove, Nicolle Mode, David R. Montgomery, Jennifer O'Neal, Jenna Scholz, Leslie Wall and the entire FE 529 class Spring Quarter 1999 at UW.

July 1999

EXECUTIVE SUMMARY

Forested landscapes in Washington State are managed by a state forest practice management system that includes Washington forest practice rules, Watershed Analyses and landowner landscape plans. One objective of the management system is to protect aquatic resources and the forest practice rules are currently being changed to increase protection. Monitoring at the watershed scale is proposed as an essential requirement to evaluate the effectiveness of the new forest practices. The Center for Streamside Studies initially responded to a Request for Proposals from the TFW Effectiveness Monitoring and Evaluation Program to provide a *Conceptual Framework, Design Process and Program Standards for Watershed Effectiveness Monitoring*. After initial discussions it was decided that an appropriate initial step was to provide a review of the document entitled "TFW Monitoring Program Watershed-Scale Monitoring Pilot Project" (Schuett-Hames 1999).

Development of appropriate effectiveness monitoring programs at the watershed scale is an extremely difficult and ultimately controversial task. Previous efforts generally have emphasized specific activities or effects with few attempts to monitor cumulative effects in a statistically sound and defensible manner. The TFW Monitoring Program proposes to assess cumulative effects by: 1) Monitoring changes in selective input processes: Mass Wasting, Surface Erosion, Riparian LWD Recruitment, Thermal Energy and Hydrology and 2) Evaluating the response of aquatic resources to changes in input processes. This is an appropriate structure for framing the questions, although quantifying changes in input processes and responses by aquatic resources

will be difficult.

Specific questions and hypothesis presented in Schuett-Hames (1999) for individual inputs and responses are addressed in the text of this review. Several themes that appear throughout the review include:

- Reference conditions are not clearly defined. Sometimes it appears that changes will be compared to natural (unmanaged) conditions but elsewhere it seems that current post-harvest conditions will provide the benchmark. We strongly advocate reference to natural conditions. How to select appropriate sites will need to be determined.
- Sampling design needs to be developed. Methods for selecting sampling sites, number of samples, required number of samples to detect significant changes, etc. needs to be expanded. Until more specific details of sampling methodology are developed, it is difficult to predict if measurable differences are likely to be detected.
- More selective diagnostic features could be chosen. For some proposed diagnostics the natural variation is too great or the expected change is too small to provide statistically significant differences. An effort should be made to identify those diagnostic features that are most likely to be useful.
- Sampling time frame needs more consideration. For some responses five-year intervals may be appropriate. For others data should be collected more frequently and for others no significant changes are expected to occur in five years. This problem may be resolved with more careful consideration of the sampling design.
- Action levels for management responses need to be specified. It is never clear how these monitoring results will be used to support or alter forest practices. Unless there are clear management implications, monitoring efforts may be inconsequential and subsequently eliminated.

Resolving many of these issues will require the type of effort that was requested in the initial RFP.

Introduction

Most of Washington State was logged at some time during the past 150 years and some lands are going into their third rotation. According the FEMAT (1993), the state is currently 50% forested lands and about 50% of the forested lands are being managed for timber harvest. Past management activities have led to degradation of watersheds and streams and contributed to the decline in fish populations. As forest practices change to increase overall ecological integrity within watersheds and to reduce the impacts on threatened and endangered salmon, it is necessary and desirable to document changes in stream and watershed conditions that occur in response to changes in forest practices.

There have been many efforts to establish meaningful monitoring programs for the Pacific

Northwest (e.g. MacDonald et al. 1991; Hayslip 1993; Schuett-Hames et al.1994; Conquest and Ralph 1998). Monitoring programs can be designed to evaluate a variety of activities and effects (MacDonald et al. 1991). For example, one may want to identify trends, baseline or reference states, or examine the effectiveness of actions. Monitoring is also used to evaluate implementation, project, and compliance activities.

The proposed TFW project seeks to document whether the forest management system (forest practices rules, watershed analysis and landowner landscape plans) used in Washington is effective in protecting aquatic resources from cumulative impacts of forest practices on watershed scale (Schuett-Hames, 1999). The scale for evaluation is Watershed Analysis Units (WAUs). Those watersheds selected evaluation will be stratified by physiographic region, geologic groups, and past management history (Schuett-Hames et al. 1999).

There is relatively little published work on how to monitor cumulative effects in a statistically sound and defensible way. The U.S. Environmental Protection agency sought to develop techniques via the Environmental Monitoring and Assessment Program (EMAP) in the early 1990s (e.g. Larsen et al. 1991) but the EMAP strategy is still being refined and evaluated. There is still a huge need for sound monitoring that can quantitatively answer questions such as those posed by the three objectives of the TFW project.

The proposed project attempts to develop a watershed cumulative effectiveness monitoring strategy for the state of Washington. The three questions stated as the primary objectives are important questions that need answering. This document provides a critical review of the proposed monitoring project to address the watershed-scale effectiveness monitoring questions in Schuett-Hames (1999).

The proposed project has the following overall objectives phrased as three questions regarding the impact of forest practices at the watershed scale.

1. How do forests practices alter watershed conditions, particularly input processes of water, wood, and sediment that affect aquatic habitat?
2. How do those changes in watershed condition/processes affect aquatic habitats?
3. Can we successfully identify those areas where forest practices are most likely to produce deleterious impacts to aquatic resources?

Schuett-Hames (1999) is structured in terms of changes in individual input processes and aquatic resource responses to those changes. Each of these is presented as specific monitoring questions and hypotheses. We comment on monitoring hypotheses and techniques for each of the three main objectives. However, we first provide general comments, provided independently by different reviewers, on the document as a whole.

General Review Comments

Objectives

The proposed task is a large and complex one that has no identifiable successful precedent. The three main objectives are important and reasonable questions. Indeed they are questions that researchers have been wrestling with for decades. To answer them will require a monitoring effort that is far more substantial and sustained, or more precisely defined, than past efforts.

Schuett-Hames (1999) proposes to evaluate changes relative to three different perspectives: current conditions, performance standards, and natural conditions. This is an admirable goal and may provide useful insights to evaluate new practices if they can be implemented. However, each of these conditions has serious limitations. Comparing new practices to current conditions that have been significantly degraded by past management conditions implies that any measurable improvement from the past is acceptable. This is a dangerous standard. New practices should certainly exceed performance standards since performance standards establish minimum acceptable conditions. If practices do not satisfy performance standard it is certainly a cause for concern. The performance standard itself, however, may provide a relatively low standard. It is not possible to evaluate these standards since many have not yet been proposed. Performance standards also do not provide any information about natural conditions. We believe "natural conditions" provide the appropriate standard. It may be difficult to establish a baseline for natural conditions and processes because so many watersheds have been disturbed previously. If a particular watershed has already been exposed to past management conditions, how can we assess natural conditions? Nevertheless, in order to gauge the effect of human activity on ecological systems change from an undisturbed condition is the appropriate measure.

The questions, as phrased in Schuett-Hames (1999), may be somewhat confusing. For example, "cumulative effects of forest practices on a watershed scale" does not allow us to isolate impacts from individual practices. We recognize that there will be independent monitoring to assess impacts of individual practices. It is unclear, however, how that information will be integrated with the watershed monitoring to quantify the effects of different practices on aquatic resources at a watershed scale. Since different practices will occur in watersheds with different history and physiography, various combinations and magnitudes of impacts from different forest practices could yield similar results or vice versa.

The current state of watersheds and the causes for their current state needs to be acknowledged more clearly. Many activities have occurred over the landscape in the last 150 years and it will be difficult to differentiate between the effects of past, present and future timber management practices. Presumably all forest practice regulations, since their inception in the early 1970's and through subsequent changes, have been formulated to protect aquatic resources. New practices will be implemented on a landscape that has, and continues to reveal, cumulative lag effects from past timber harvesting.

There is heavy reliance on information collected during Watershed Analysis. However, Watershed Analysis has some limitations (Collins and Pess 1997). Watershed Analysis conducted in different watersheds by different analysts may not provide for interpretations that are consistent enough to be lumped with other Watershed Analyses into a large enough data base for further analyses. It is also not clear at this time if the Landowner Landscape Plans will be similar enough to Watershed Analysis to be included in such efforts. In addition some Watershed Analyses will not provide sufficient data for monitoring purposes since a standard method for all assessments has not been established (e.g. WSA Fisheries Module survey methods can use any established means for collecting data).

Watershed Analysis may provide information with which to compare current forest practices to past forest practices, but it does not provide the information necessary for comparing current practices to "natural" or background disturbances. To do this, the "natural conditions" and effects of "past management" need to be known. This requires more than simply monitoring; it requires substantial knowledge of the landscape that extends well beyond the information typically collected in a WA Watershed Analysis. There is little to no mention in the project document that addresses how "natural" conditions will be estimated. This is critical to a successful monitoring program and needs to be elaborated on.

The question on identifying "sensitive" areas will be difficult to answer because not all "sensitive areas" will necessarily be impacted, and some "non-sensitive" areas may be affected. More importantly, forest practices may be modified in "sensitive" areas to reduce the probability of adverse changes that affect aquatic resources. If the practices are altered on identified sensitive areas, then "sensitive" and "non-sensitive" areas will be exposed to different forcing functions. This leads to many complications in quantitative analysis of monitoring data. This question seems to be more appropriate to an evaluation of watershed analysis, and perhaps should be addressed in an independent study plan.

Statistical issues

If one of the main goals of this project is to develop a data set that can be used to identify cause and effect relationships (Schuett-Hames et al. 1999), then sampling methods, criteria to identify changes and analysis methods need more development.

The first part of any sampling or monitoring design is defining the objectives. Detailed objectives are useful in creating sampling designs and in prioritizing potentially competing concepts, (e.g. status vs. trends). A step that is often over looked, is defining the answer. Is it only after the answer is defined that the researcher can determine if the study has the possibility of meeting the objectives. For example, if the question is "Have there been any significant biological changes to stream X since the nearby harvesting of the forest?" the answer will need to include a measure of change since baseline, for variables that quantify "biological change", and "significant" will need to be defined statistically and/or biologically. Related to this is the problem of what will be compared. At times it appears that the comparisons will be present versus past and present versus future (trends). At other times it appears that managed systems will be compared with natural, or unmanaged systems. At still other times it appears the comparisons will be only with standards set out by the forest practices board. Clarification is needed on how "natural", background or target levels will be determined.

In the proposed project the criteria for determining a significant change are not defined. Numerous references are made to various properties "increasing over time", or "decreasing", or "being less than they were under past practices" but no statistical basis for those comparisons are described. Given the enormous natural variation in most of the proposed variables, the large errors in quantifying them at watershed scales, and the uncertainties about relevant time scales for monitoring responses, establishing acceptable statistical criteria will require substantial effort.

The scope or scale of sampling will determine the types of effects that can detected. Sampling at a large scale, such as watersheds or ecosystems, allows for detection of effects that may not be observed when monitoring a single stream, or river reach. General equilibrium effects appear only at the aggregate level, and one species may not indicate much about the overall stability on a small scale or in the laboratory. However, monitoring at the watershed scale may require different metrics than are used at the impact site to capture the cumulative effects information that is being sought.

Processes outside the watershed (e.g. climate changes, ocean temperatures) that affect stream and watershed ecology create challenges to inference when looking for the impacts of forest practices. These effects confound inference to other areas. Other confounding factors include effects from activities upstream that may impact the study area and effects from downstream such as dams, ocean harvest of fish or introduction of exotic species. These also increase the difficulty of isolating the effects of forest practices.

Baseline data must be included in any monitoring program. The appropriate baseline to use for this project will depend upon the perspective chosen: current conditions, natural conditions or performance standards. Frequency of sampling is also important. The US Forest Service Forest Inventory Analysis (FIA), and the Forest Health Monitoring Program (FHM) revisited sites at 4 to 10 year intervals (Olsen et al. 1999). That sampling period may be too infrequent to detect interesting changes or to identify and establish cause and effect relationships (Olsen and Schreuder 1997).

All processes within the watershed will be influenced by natural variation that occurs at yearly, decadal and longer frequencies. Some processes will need to be monitored for 50-200 years to detect changes while others will respond within a few seasons. However, monitoring is proposed for every 5 years, for an undefined duration. The time interval is too frequent for some process, and too infrequent for others. Time intervals between sampling events should be specified when the specific question and diagnostic features to be measured are better defined.

Numerous state and federal agencies are currently conducting watershed monitoring programs with considerable duplication of sampling efforts. Each agency has a unique focus, but many collect data for the same geographic areas and ask similar questions. Recently a group in Oregon examined the feasibility of combining several survey efforts into one monitoring program (House *et al*. 1998). They provide a common basis for creating one survey, and include within the design repeated visits by different teams to examine measurement repeatability. They recommend transition to a national integrated inter-agency resource status and trends inventory. The advantages and disadvantages of this approach will require extended debate that may not be consistent with TFW's time frame to establish an effectiveness monitoring program. The issue is valid, however, and more attention should be given to eventual analysis and maximum utility of these data.

The following section discusses the effectiveness questions for individual input processes and aquatic resource responses. We have tried to elaborate on changes that will improve the project's chances of success.

Effectiveness Question 1. ***How are input processes and watershed conditions responding to cumulative impacts of forest practices on a watershed level?***

Approach:
A watershed will be impacted by cumulative effects of different forest practices and by similar practices applied at different times and locations. Therefore, at the watershed scale, it is desirable to monitor the "rate and magnitude of input processes .... throughout the watershed". This is a difficult and expensive project, if it is feasible at all, and will provide limited information to quantify cause-effect relationships. It appears that the general approach will be to measure some parameters near the mouths of streams as they are exiting the watershed. If that interpretation is correct, monitoring may be able to detect changes in input processes at these downstream reaches but is unlikely to identify the cause(s) for those changes. Also, it will not detect impacts that occur in the upper watershed but are not propagated downstream. Furthermore, it will be difficult to determine if these changes are a result of "forest practices conducted under past or current forest practices, other land uses or natural events".

An alternative approach is to consider practices as they affect specific stream reaches and then calculate watershed impacts by scaling up for different types of terrain. This approach is used for WA Watershed Analysis and seems to have been proposed for Effectiveness Question 2. Reviewers did not reach consensus on the most appropriate approach. It will depend upon the final objectives of the monitoring effectiveness program and the importance of establishing cause-effect relationships.

MASS WASTING

Monitoring Question: ***How has sediment delivery to streams from mass wasting changed over time in response to forest practices conducted under the state forest practice management system?***

Three hypotheses are presented to address this question: 1) Rate of mass wasting and volume of sediment delivered to stream will decline, 2) Sediment delivery will meet performance standards, and 3) Precipitation and soil moisture influence mass wasting. The proposed monitoring variables are mass wasting rate (acres/event/year) and sediment delivery to streams.

This question is too vague, and doesn't seem to recognize the lag effect associated with harvest and subsequent hillslope failures. However, the approach of trying to identify the background levels of sediment input, attributing volumes of sediments to specific sources (natural vs. management) and evaluating whether the "new" practices are producing less sediment input is sound in concept. Such an approach currently lacks a sound analytical framework, although one approach is provided in the Mass Wasting Module from the WSA methods. Alternative techniques are being used to develop sediment TMDL's for Idaho and for the Simpson HCP/TMDL in Washington (Fitzgerald *et al*. 1998). The success of this approach depends upon a solid system to account for the spatial distribution of unstable land forms, past and potential sources, what specific sediment abatement actions are taken, and how these relate to demonstrable sediment reductions in the future.

Use of the metrics, changes in rate of landslides and delivery of sediment from mass wasting events, is a vague and weak standard. In fact, taken literally it would mean that if future rates of landslides do not exceed the worst observed under past forest practices then no change in management style is required. This is a recipe for institutionalizing the acceptability of high rates of landsliding. A reasonable standard needs to at least define how much less landsliding than observed under past practices is acceptable and a more reasonable standard would be developed relative to natural background rates with a specified time window within which some variability is tolerated due to extreme events.

Performance standards of no sediment from new roads and harvest units is an unrealistic standard because there will be some sediment delivery from roads and from harvest units. They key issue is how much is delivered. One could use a standard of no measurable increase above background rates, but that would require information on background rates of sediment delivery under current and natural conditions. Those data do not exist for most watersheds.

The last hypothesis is not really an hypothesis and should be removed. "Precipitation and soil moisture will influence mass wasting frequency and magnitude", as will slope, soil type and time since harvest. We know this to be true. The point that we must interpret any observed landsliding within the context of the recurrence interval of the storm and antecedent moisture conditions is of course correct, but this does not form a testable hypothesis.

Another hypothesis should be added to this list: Forest practices will not increase the rate of mass wasting over unmanaged rates.

Using the metric (acres/event/year) is more difficult than #/year and does not obviously provide more valuable information.

Sediment delivery is extremely difficult to obtain and is not available as baseline condition for most watersheds.

Monitoring and Evaluation:
In this section there seems to be ambiguity as to whether we are looking for convergence with "natural" rates, or just improvement from some previous value. That may be an unavoidable consequence of attempting to address multiple hypotheses simultaneously. We believe the emphasis should be placed on comparison to natural conditions. "Trends in natural and management induced mass wasting will be evaluated in the context of storm event magnitude and frequency" implies that there will be unharvested (natural) watersheds to serve as controls. That seems unlikely and those watersheds will not be comparable to managed watersheds. For example, we cannot determine the effect of a particular cutting practice by comparing it to a natural watershed that does not have roads.

Use of existing WA mass wasting inventories as the "baseline for past forest practices" is not consistent with the stated goal of using some measure of deviation from natural background rates as a metric for evaluating forest management performance. Past WSA are variable in quality. It is not valid to assume that they will provide reliable information on past forest practices. More importantly, using past rates as a benchmark essentially defines poor management practices of the past as the acceptable standard of the future.

It is frequently unclear what time frame is represented by an initial inventory. If we conduct inventory at regular intervals and accurately record and compare events, we can calculate rates of mass wasting events between sampling times. However, the initial survey will represent a longer and unknown time interval. Thus, it will provide an inflated value for baseline conditions. This baseline will represent past forest practices but not natural conditions. Five years is not an appropriate time frame to evaluate changes in the frequency of mass wasting events. Sampling at five year intervals will be confounded by past forest practices. Since mass wasting events tend to peak 10-15 years after harvest, it may be difficult to define the time period being evaluated. Because of the lag between harvest and slides it will take a long time to evaluate effects of past practices, and we may be attempting to evaluate different practices with the same data points.

Also, with five year sampling, it will take 20 years to obtain 4 data points and that may not be sufficient to establish a statistically significant trend.

A more rational way to assess performance of future forest practices is to determine background rates with a measure of acceptable variation around those rates. This would provide the benchmark rather than current conditions that have led to the need to adopt some sort of monitoring program.

It is unclear how the information generated by mass wasting monitoring will be used. We expect that results of the monitoring program will influence future decisions on how to harvest different types of slopes. However, no mechanism is presented for how the monitoring results will be used to modify, shape, or retain future management practices. If there is no such feedback, there is no point in monitoring.

SURFACE EROSION

Question: ***How does sediment delivery to streams from surface erosion change over time in response to forest practices conducted under the state forest practice management system?***

The Hypotheses presented to evaluate this question are: 1) Volume of sediment delivery from soil erosion will decline, relative to past forest practices, in response to new forest practices, and 2) Volume of sediment delivery will be below performance standards. The second hypothesis appears to relate to mass wasting rather than surface erosion. The proposed monitoring variable as for mass wasting is volume of sediment delivered to stream. This parameter will present the same measurement difficulties as described above for sediment delivery from mass wasting.

Many of the comments presented above for Mass Wasting will also apply to this question. In fact it is difficult to distinguish between sediment introduced into streams by mass wasting or by soil erosion. No methods are presented for making measurements.

Updating the sediment budget at 5 year intervals will do little or nothing to document changes in sediment inputs associated with the new forest practices. This can only be documented by a specific monitoring program that measures surface runoff volumes before and after, and links this with specific sediment abatement actions at sites selected on the basis of a strategic sampling design (Rashin et al. 1999). Total sediment volume may not be the most appropriate measure for soil erosion, particle size distribution or delivery timing may be more important than total volume and those will not be measurable at five year sampling intervals. The best estimates of basin sediment budgets have error terms of 50%, so documenting sediment reductions attributable to specific land management BMP's will need to be carefully designed and executed.

If volume delivered is a function of production, it may be more efficient to evaluate production as a surrogate measure for sediment delivery.

Monitoring and Evaluation:

This does not appear to be a desirable process to include in the effectiveness monitoring program. It is unlikely that surface erosion can be measured accurately enough to detect changes over time. And we cannot distinguish between natural and management-induced sediment; we can only measure differences in total. The results depend upon accuracy of the "partial sediment budget" approach in WA watershed analysis but this approach has not been validated and/or peer reviewed outside of the TFW process.

RIPARIAN LWD RECRUITMENT

Question: ***How has LWD recruitment potential of riparian zones changed over time in response to forest practices conducted under the forest practice management system?***

This is an appropriate question but probably cannot be answered on a five-year time scale. Percent recruitment potential may not change measurably during a five-year interval even if the condition of the riparian zone is improving. It would be preferable to provide more detailed information on stand characteristics/stand dynamics. Although the trend ("will increase over time") in LWD recruitment is important, the rate at which it changes is also important for gauging stream recovery.

The Hypotheses proposed to test the question are: 1) The number of stream miles with high short term LWD recruitment potential will increase over time as the current forest practice management system is implemented, and 2) Riparian Stand will meet performance standards of the forest and fish report.

The value of the first hypothesis is not evident to most reviewers. This hypothesis assumes that the quality of riparian zones will improve with time as the new forest practices management system is developed. The validity of that assumption depends upon adequate prescriptions that are consistently enforced.

The monitoring variables proposed to test these hypotheses are short term recruitment potential of riparian stands and percentage of recruitment potential. It would be useful to know if potential recruitment actually becomes recruited. Short term recruitment may result from windthrow that adversely effects long-term recruitment. There needs to be discussion of how recruitment potential will be determined with more details on sampling methodology. Will high resolution photography be required for the entire basin? How will watershed scale data be developed?

How will retention of a minimum riparian zone coupled with an adjacent managed zone (which is called for in the new regulations) actually "increase over time the number of stream miles with high (?) short term LWD recruitment potential"? A better hypothesis might be that LWD recruitment from riparian zones subject to adjacent timber harvest, have the same LWD input rates as those stream reaches where no management has occurred. Also, could add that integrity and forest characteristics (e.g. height, diameter, stem density, wind throw and mortality rates) are not altered by forest practices or prescriptions.

The Monitoring and Evaluation Procedures (Schuett-Hames et al 1999) state that recruitment potential is based on 1993 date that were collected following years of past management. Comparisons with these data may provide evidence of change from current conditions, but will not adequately evaluate the new system. Recruitment potential should be determined for natural stands so the new-forest management system can be compared to desired future conditions of *natural* potential.

Performance standards for the eastside of the state are not provided.

How sensitive is percent recruitment potential to changes in stand characteristics/stand dynamics within five-year periods?

Monitoring and Evaluation:
If the number of miles with "high short term recruitment potential" can be determined accurately, it would provide a single data point for each watershed every five years. That value may not change in some intervals even if the quality of the riparian zone is improving. The assumption that the new prescriptions will better protect existing conifer dominated stands on fish bearing waters needs to be placed in context - better protection than what, no harvest at all or better than past clear cuts? Any alteration of the stream side and hillslope stand composition will not improve the LWD recruitment potential over what it is now, unless you believe that the

increased blowdown will be positive improvement. Currently, LWD levels within streams where riparian zones have been harvested are generally lacking in wood, and the riparian stand composition is inadequate to provide LWD in sufficient size and quantity to replenish the supply in the near term (next 20-50 years). We do not know the present vegetative condition of riparian stands along streams on private forest lands, so it is hard to say what the potential LWD recruitment might be. It is also important to understand that different mechanisms account for different LWD input rates and characteristics. Streams with a confined morphology (low ratio of channel width to valley width) receive their wood input from episodic hillslope failures (which makes a case for not cutting those adjacent hillslopes) while other streams may receive most LWD from blowdown or undercut bank processes. Recognition of these different mechanisms is very important if one is to construct a credible approach to recruitment potential.

THERMAL ENERGY (RIPARIAN SHADE)

Question: ***How has the shade provided by riparian stands changed over time in response to forest practices conducted under the state forest management system?***

This is a reasonable question although it may be difficult to answer. A more appropriate question might address stream temperatures directly rather than shade, which is difficult to measure. In addition the quality of shade may vary in a manner that affects the stream but cannot easily be quantified.

The Hypothesis proposed to test this question: The number of stream miles with riparian stands that meet the target shade conditions in the forest practices rules will increase over time as forest practices are conducted under the state forest practices management system, is confusing. It seems likely that some unknown percentage of streams on private forest lands have riparian stands with mature conifers that provide adequate shade (the mature group), while some other unknown number of stream miles do not (the recovery group). The recovery group will need time to re-grow to provide the needed shade. Time lines for this to happen will be highly variable. When this group is harvested, we expect that the riparian prescriptions will retain all or most of the effective shade at these sites, thus protecting the water temperature profile characteristic of these stream-riparian complexes. However, how will the mature group, those stands that will be harvested in the next decade, have their effective shade levels "increased" by being included in a harvest unit? Hopefully, the shade levels will only decrease slightly with no measurable net increase in stream temperatures. It is this group that should be subject to the test of effectiveness of the riparian prescriptions before and after harvest occurs.

This is not a statistically testable hypothesis. The result of the analysis will be a single number (with no measure of central tendency) at five-year intervals. However, one could evaluate % of watershed that is shaded for a number of different watersheds and get some measure of change. It is also a relatively insensitive measure that is unable to detect improving conditions once the shade targets are met. For example, if shade targets are met with deciduous trees or small conifers and do not change as the riparian stand ages, no improvement will be apparent even though the quality of the riparian zone is improving over time.

There are no performance standards stated in the hypotheses. As for other input processes we believe the most appropriate comparison is with natural conditions.

The monitoring variable to be used to evaluate recruitment potential is percent of obscured stream channel determined from aerial photos. It is unclear how accurately stream coverage can be determined from air photos. It seems likely that the evaluation will vary with season, tree species and height.

It seems preferable to measure thermal changes directly. Put a thermograph into the stream of interest before harvest and use the temperature data to determine change directly. Also, install air temperature monitors to measure potential changes in riparian microclimate. Some other related parameters that may be affected by changes in riparian vegetation removal include relative humidity, wind speed, hydrology, and ground temperature.

Monitoring and Evaluation:

The procedures should be rewritten to specify that data will be collected on the initial number of stream miles that meet riparian shade targets. Data collected at later times will be used to develop a trend that indicates if an increase or decrease in total stream miles meeting shade targets has occurred as a result of adhering to the forest practice rules. It is unclear how natural conditions will be evaluated and those will change for different valley types. What controls will be used at the watershed level - adjacent watersheds not using forest practices rules?

How can the riparian canopy closure assessment module of the WSA procedure be used as a baseline for management if past practices were inadequate to provide riparian shade?

Rather than use an initial aerial photo assessment of riparian shade, it would be more useful to conduct an historic aerial photo analysis to determine what the "natural" riparian condition was.

Five-year intervals seem inappropriate given the rate of riparian vegetation growth. It may be adequate for deciduous vegetation re-growth, but conifer shade will not be adequately addressed. Also, this is an excellent area to use thermographs and monitor direct changes in thermal conditions. These changes are quickly and easily apparent and a five-year review of the recording data will yield results useful in determining the effectiveness of forest practices.

HYDROLOGY

No hypothesis is presented for hydrology so it is not possible to evaluate this section. It will be particularly important to have natural watersheds, or long term data sets for comparison, since hydrology will vary markedly on fluctuations in precipitation and temperature.

Effectiveness Question 2: ***How do stream channels and aquatic resources respond to changes in input processes altered by forest practices?***

Approach:
The attempt to measure changes in channels and aquatic resources due to changes in forest practices is a shift from identifying changes in processes to looking for symptoms of those changes. Historically the emphasis in monitoring has been on symptoms and it has never told us how to fix the symptoms or what the cause(s) of the symptom were. Obviously we want to know how aquatic resources are affected by changes in controlling processes that may be altered by forest practices. However, we are not aware of an agreed upon set of appropriate diagnostic features and the problem of interpretation is compounded by legacy effects, inherent variability, natural disturbance, etc. Many diagnostic features (e.g. pool area, pebble counts, bank flow width, habitat units, LWD counts) have been shown to exhibit poor repeatability in measurement among different field personnel. This creates an insurmountable problem for statistical analysis.

The diagnostic features that are presented are all structural. The emphasis under Question 1 was on input processes, but there is no mention of any aquatic processes under Question 2. The inclusion of some biological diagnostics may get closer to identifying changes within the system that significantly affect aquatic resources in basic and critical ways.

The approach for assessing aquatic resources is different than that for assessing changes to input processes. The proposed approach appears to be geared for selecting reaches within a watershed for sampling. The general idea seems to be to measure responses (symptoms) in selected reaches and then attempt to rescale to watershed scale.

The approach of using dominant "response situations" as the primary monitoring approach has some drawbacks. First, the approach may completely miss the signal of land use impacts if the selected reaches are not those impacted by particular events. The example of a large landslide occurring in the watershed provides cause for worry. A single large landslide can dominate the sediment budget of even moderate sized channels, but it may or may not occur in the right place to influence preselected reaches of "response situations". In addition the idea that sampling should occur only once every 5 years is a sure fire way to make sure that no trends are discovered for at least 15 years as two points don't make a trend. Data for channel or landscape attributes that are expected to exhibit substantial variability should be sampled annually in order to assess trends and to build a reasonable sample size at each place to offset the inherent variability between observers. Sampling every 5 years is unlikely to detect meaningful trends and may obfuscate relations and delay implementation of needed changes in management until a decade after symptoms are expressed in the landscape.

The diagnostics used under this question are predominantly physical parameters even though aquatic resources was defined earlier in the document to include populations of organisms. Some diagnostic features mentioned might be relevant, but it will be difficult to differentiate between background and management-induced sediment already in the channel that is migrating downstream vs. new sources resulting from management actions. This analysis could be quite powerful if it were spatially and temporally linked with a more developed examination of what new sources of sediment are being triggered from natural and management related actions. The sampling scheme should be better defined, since significant errors can occur when attempting to characterize whole reach sediment characteristics with only site sampling. The time frame within which one could expect to see results needs further refinement. Many authors have wrestled with monitoring this important aspect of fish habitat (see multiple sources cited in Klein 1997, Chapman 1988, MacDonald 1997).

It is necessary to already know something about "aquatic resources" responses to changes in input functions in order to establish hypotheses for expected responses in a watershed. What we really need to address for a given watershed is whether changes in input processes are actually influencing the characteristics that we believe are sensitive and then to determine if those changes actually influence the aquatic organisms of interest/concern. Many of the questions in this section are phrased as research questions not monitoring questions.

The list of diagnostics for assessing changes in sediment, wood or water loading is long and needs to be more thoroughly assessed for value. A recent document by Scholz and Booth (1998) assessed stream habitat protocols in urbanizing areas and developed 3 levels of assessment that would be useful in assessing stream condition. Looking at their work may help refine and decrease the number of diagnostics suggested for monitoring in this project.

As mentioned under general comments earlier in this report, it will be necessary to identify the magnitude of change for monitored variables that are of concern. Keying the assessment of the resource to the direction rather than the magnitude of change is inadequate. Obviously any management regime that does not result in improvements to habitat much be considered inadequate if the monitoring program was developed because current conditions are inadequate. Hence, the direction of change for compliance with watershed management goals is given; the magnitude of change that constitutes improvement is the real issue.

FINE SEDIMENT

Monitoring Question: ***How do aquatic resources respond to management-induced changes in fine sediment input on a watershed scale?***

This is an appropriate but very general question. The list of diagnostics is overwhelmingly abiotic although the diagnostic that is predicted to be most sensitive (Appendix Tables 1-3) is aquatic invertebrates. Many of the selected diagnostics have been previously shown to be either rather insensitive to sediment changes or to have a very high variability in both expression and measurement.

V* is a relatively insensitive measure of fine sediment loading (Lisle and Hilton 1999); it may vary by only a factor of two over three orders of magnitude change in sediment yield. Fine sediment content of the subsurface sediment in riffles (or on bars) may be a better indicator of the fine sediment content of the sediment load. Reduction in pool depth is another measure that is expected to be a good diagnostic for increase sediment supply (Collins et al. 1995) although it may not distinguish between fine and coarse sediment.

For any diagnostic, methods need to be established for data collection that are repeatable among field technicians (Poole et al. 1997) and studies should be completed to confirm that the chosen diagnostic(s) actually are diagnostic. Work is especially needed to identify useful biological measures that can be more closely associated with the health of fish populations (Karr 1991). Additional research with aquatic invertebrates is required to determine the most effective parameters to use.

COARSE SEDIMENT

Monitoring Question: ***How do aquatic resources respond to management-induced changes in coarse sediment input on a watershed scale?***

Similar comments apply here as above for fine sediment. Changes in the input of coarse sediments are most likely to be apparent in physical rather than biological diagnostics.

LWD RECRUITMENT

Monitoring Question: ***How do aquatic resources respond to management-induced changes in the LWD recruitment on a watershed scale?***

The predicted responses to changes in LWD recruitment account for only the effects of LWD on physical structures and pool dwelling salmonids. Wood has other functions including dissipating energy during high flow events and increasing production of macroinvertebrates. This latter observation is in direct contrast to the predictions in Tables 1-3. Increased production may have important ramifications for food supply of salmonids and other fish. An adequate evaluation of habitat and ecological conditions should include biological monitoring of ecological health of the system (as measured by fish and invertebrates) which could be compared to LWD loading levels to determine correlations between LWD and various measures of ecological integrity.

Biological assessments of fish should concentrate on resident rather than anadromous species.

As with coarse and fine sediments, we already know something about "aquatic resources" responses to changes in these input processes. However, we need to know if changes in the inputs actually elicit a response for a given watershed. This is a research rather than a monitoring question.

This section assumes that potential recruitment measured in riparian zones for effectiveness question 1, actually appears as LWD in the stream channel and has benefits for aquatic resources. The time frame for seeing those benefits is unknown but certainly longer than five years.

THERMAL ENERGY (Riparian Shade)

Monitoring Question: ***How do aquatic resources respond to management induced changes in riparian shade on a watershed scale?***

This question assumes that riparian shade is an appropriate measure of changes in thermal input and limits diagnostic features to maximum temperature. Although this may be a useful parameter for defining changes, the maxima is not as biologically important as other metrics including the daily mean, maximum weekly average temperature (MWAT), and the 7-day rolling average of the mean (Sullivan 1990). Also, the magnitude, frequency, duration, timing, rate of change, variation, and spatial extent are important monitoring variables to consider and need to be considered when conducting a trends analysis.

With respect to the hypothesis given, air temperature should be considered in addition to elevation.

It may be more important to evaluate biological responses within the watershed. We propose that temperature, as recorded by remote thermographs, be considered as the input variable and biological responses be considered as the "aquatic resource".

HYDROLOGY

Monitoring Question: ***How do aquatic resources respond to management-induced changes in hydrology on a watershed scale?***

As for previous sections, the monitoring question addresses aquatic resources but the diagnostic parameters concentrate on physical characteristics of the stream. Some attempt should be made to couple changes in hydrology with observed biological responses. Again, this may initially require a research effort but organisms are the ultimate "aquatic resource". Recent observations in Oregon, and locally in the Cedar River suggest that high flow events scour out populations of predatory fish, especially sculpin, and promote increased survival of juvenile anadromous fish. Those effects cannot be observed by monitoring for physical characteristics only.

Monitoring and Evaluation:
Some concerns about monitoring and evaluation are addressed in our general discussion on the approach to monitoring effectiveness question 2. Although "response situations" provide an approach for obtaining watershed scale effects, selection of the "correct" sites is problematic. Selection of enough sites to characterize different sets of conditions within the watershed and provide some measure of variability for different types of "response situations" will be

necessary. It is not sufficient to predict the range of expected values for diagnostic features; it is also necessary to include a measure of variability within that range. Without a measure of variability, recovery efforts are likely to congregate around minimal acceptable conditions. We are unaware of readily available compilations of expected values although the data may be available to compile that information for some diagnostic features.

A particularly challenging aspect of this problem is indicated by the example of the "response situation" provided: high input of fine and coarse sediment, low LWD and low peak discharge. For most diagnostic features some changes in input processes cause the diagnostic feature to increase while changes in other input processes cause the diagnostic feature to decrease. Thus, any observed changes are "expected" under some channel response hypothesis. We do not believe data are currently available to predict rates of change for most diagnostic features.

Effectiveness Question 3: ***Does the Forest Practice Management System Successfully Identify Sensitive Locations?***

The mandate for information collected in this section is very loose. If there is no strong direct feedback between monitoring and management, then the incentive to conduct monitoring will diminish. Perhaps there should be mandatory changes in land management if monitoring determines that sensitive areas are not identified. Again this is a question that is more appropriate to address in a review of Watershed Analysis than Effectiveness Monitoring.

SAMPLING DESIGN

The sampling design is a critical, but not yet developed, component of this monitoring program. It appears that mass wasting, surface erosion budgets, LWD recruitment potential, temperature, and hydrology will be determined for every watershed in which more than 50% of the land is managed under the forest practices management system. That is an enormous and expensive task which may be addressed in a more cost-effective manner by stratification and sub-sampling. However, an original objective of the WA Watershed Analysis methodology was to evaluate all watersheds independently within their unique spatial context and history of disturbance. Channel responses to land management decisions will reflect the unique characteristics and history of individual watersheds. There is no reason to expect selection of "representative" watersheds to in fact be representative. A credible monitoring program should include all intensively managed watersheds.

The overwhelming number of very general hypotheses described in the text and appendices, make it difficult to determine what is actually proposed. And there needs to be some guidance regarding the methodology of selecting response reaches. If all response reaches are evaluated and the evaluation methodology is sound, then the approach may be appropriate. If, however, few reaches are sampled with poor protocols, the results will not be valuable. The proposal cannot be evaluated adequately without specifying the actual monitoring protocols.

CONCLUSIONS

Schuett-Hames (1999) and Schuett-Hames et al (1999) have identified important issues that need to be addressed before the effects of new forest practices on Washington state watersheds can be determined. The general approach: 1) To determine the response of input processes to "forest practices conducted under the state forest practice management system" and 2) To determine how "aquatic resources respond to management-induced changes in input processes on a watershed scale" is credible and presents a laudable goal. We cannot expect, however, to satisfy that goal for most of the diagnostic features that were selected due to large variation that occurs for many variables and relatively modest changes that can be expected from new forest practices. In order to develop a cost-effective watershed effectiveness monitoring program, it will be necessary to quantify our current understanding of processes that occur in natural and managed watersheds and select diagnostic features with high signal/noise ratios.

REFERENCES

Barbour, M.T., J.B. Stribling, J. Gerritsen, and J.R. Karr. 1996. "Biological Criteria: Technical Guidance for Streams and Small Rivers." EPA 822-B-96-001. Revised ed. United States Environmental Protection AgenCy.

Chapman, D.W. 1988. Critical review of variables used to define effects of fines in redds of large salmonids. Trans. Am. Fish Soc. 117: 1-21.

Cochran, W.G. 1977. *Sampling Techniques*. New York: John Wiley & Sons.

Collins, B.D. and G.R. Pess, 1997. Evaluation of forest practices prescriptions from Washington's watershed analysis program. J. Am. Water Res. Assoc. 33: 969-996.

Conquest, L.L., and S.C. Ralph. 1998. "Statistical design and analysis considerations for monitoring and assessment." Pp. 455-75 in *River Ecology and Management*, eds R. J. Naiman and R. E. Bilby. New York: Springer-Verlag.

Conquest, L.L., S.C. Ralph, and R.J. Naiman. 1994. "Implementation of large-scale stream monitoring efforts: sampling design and data analysis issues." Pp. 69-90 in *Biological Monitoring of Aquatic Systems*, eds S.L. Loeb and A. Spacie. Boca Raton, Florida: CRC Press, Inc.

Everest, F.H. and D.W. Chapman. 1972. Habitat selection and spatial interaction of juvenile chinook salmon and steelhead trout in the South Aloutte River. Prov. B. C. Fish. Tech. Circ. No. 32:19p.

Fausch, K.D. 1984. Profitable stream positions for salmonids: Relating specific growth rate to net energy gain. Can. J. Zoology. 62:441-451.

Forest Ecosystem Management Team (FEMAT). 1993. Forest ecosystem assessment: an ecological, economic, and social assessment. Report of the Forest Ecosytem Management Team to President Clinton. USDA Forest Service, Portland, Oregon.

Fitzgerald, J., T. Hardy and T. Geierldaho. 1998. Refinement of a geomorphic risk assessment and application to a nonpoint sediment source analysis in the Middle Fork Payette River subbasin, Idaho. Draft manuscript prepared for the U.S. Environmental Protection Agency, Idaho Operations Office, Boise.

Green, R.H. 1979. *Sampling Design and Statistical Methods for Environmental Biologists*. New York: John Wiley & Sons.

Hayslip, G.A. (ed.) 1993. EPA Region 10 in-stream biological monitoring handbook (for wadable streams in the Pacific Northwest). EPA/910/9-92-018.

Hilborn, R., C.J. Walters, and D.Ludwig. 1995. "Sustainable exploitation of renewable resources." *Annual Review of Ecology and Systematics,* 26:45-67.

House, C.C., J.J.Goebel, H.T.Schreuder, P.H. Geissler, W.R. Williams, and A.R. Olsen. 1998. "Prototyping a vision for inter-agency terrestrial inventory and monotoring: a sustainable perspective." *Environmental Monitoring and Assessment*, 51:451-63.

Karr, J. 1981. "Assessment of biotic integrity using fish communities." *Fisheries* 6(6):21-27.

Karr, J.R. 1991. Biological integrity: a long-neglected aspect of water resource management. Ecological Applications 1:66-84.

Kaur, A., G.P. Patil, S.J. Shirk, and C. Taillie. 1996. "Environmental sampling with a concomitant variable: a comparison between ranked set sampling and stratified simple random sampling." Journal of Applied Statistics 23: 231-55.

Larsen, D.P., S.J. Christie, S.G. Paulsen, R.M. Hughes, and C.B. Johnson. 1991. EMAP-surface waters 1991 pilot report. EPA/620/R-93/003, Corvallis Environmental Research Lab.

Lisle, T.E., and Hilton, S., 1999, Fine bed material in pools of natural gravel bed channels. Water Resources Research 35: 1291-1304.

MacDonald, L.H., A.W. Smart, and R.C. Wissmar. 1991. Monitoring guidelines to evaluate the effects of forestry activities on streams in the Pacific Northwest. EPA/910/9-91-001.

Mundie, J.H. 1969. Ecological implication of the diet of juvenile coho salmon in streams, p 135-152 in T.T. Northcote (ed.) Symposium on salmon and trout in streams. H.R. MacMillan Lectures in Fisheries, University of British Columbia, Vancouver, B.C.

Murtaugh, P.A. 1996. "The statistical evaluation of ecological indicators." *Ecological Applications* 6:132-39.

Olsen, A.R., and H.T. Schreuder. 1997. "Perspectives on large-scale natural resource surveys when cause-effect is a potential issue". *Environmental and Ecological Statistics* 4:167-80.

Olsen, A.R., J. Sedransk, D. Edwards, C.A. Gotway, W. Liggett, S.L. Rathburn, K.H. Reckhow, and L.J. Young. 1999. "Statistical issues for monitoring ecological and natural resources in the united states." *Environmental Monitoring and Assessment* 54:1-45.

Rashin, E., C. Clishe, A. Loch and J. Bell. 1999. Effectiveness of forest road and timber harvest best management practices with respect to sediment-related water quality impacts. Washington Dept. of Ecology. TFW-WQ6-99-001. Olympia, WA

Scholz, J.G. and D.B. Booth. 1998. Stream Habitat Assessment Protocols: An Evaluation of Urbanizing Watersheds in the Puget Sound Lowlands. Center for Urban Water Resources Management, University of Washington. 21 pp.

Schuett-Hames, D. 1999. TFW Monitoring Program: Watershed-Scale Monitoring Pilot Project (Draft). Northwest Indian Fisheries Commission.

Schuett-Hames, D., K. Lautz, J. Light, R. McIntosh, D. Smith, N. Sturhan, K. Sullivan, and G. Wilhere. 1999. TFW Effectiveness Monitoring and Evaluation Program Plan.

Schuett-Hames, D., A. Pleus, L. Bullchild and S. Hall. 1994. TFW Ambient Monitoring Program Manual. TFW-AM9-04-001.

Skalski, J. R. 1990. "A design for long-term status and trends monitoring." *Journal of Environmental Management* 30:139-44.

Stewart-Oaten, A. 1996. "Goals in environmental monitoring." Pp. 17-28 in *Detecting Ecological Impacts: Concepts and Applications in Coastal Habitats*, eds R. J. Schmitt and C. W. Osenberg. San Diego: Academic Press, Inc.

Sullivan, K. 1990. The physics of stream heating: An analysis of temperature patterns in stream environments based on physical principals and field data. Weyerhaeuser Technical Report 044-5002/90/1.

Thompson, S. K. 1992. *Sampling*. New York: John Wiley & Sons.