The Peer Review Panel met for the third and final time in Portland, OR on November 8, 2001. At this meeting, the panel discussed and reviewed the overall approach the Workgroup has developed for establishing temperature criteria. Specifically we reviewed two documents (Appendix A and Appendix B), which form the basis for EPA's Proposed Approach to Development of Temperature Criteria Guidance.

Below are the important points of that meeting, presented in 2 sections. In the first section, I summarize the discussion from the meeting as well as from follow-up email communication and include ideas that the Peer Review Panel felt were important to mention, but that did not show up in the questions. The second section contains our responses to the specific questions put forward by the Technical Working Group.

Sherri L. Johnson, Ph.D.
Peer Review Panel Lead Scientist

Section 1.

Summary comments from Peer Review Panel discussion November 8, 2001 regarding Draft of EPA Region 10 Stream Temperature Criteria

The Peer Review Panel recognizes and appreciates the thought, effort and detail that have gone into these revisions of the Stream Temperature Criteria. The panel agreed that the ultimate goal of the criteria is to remove stream temperature as a limiting factor or impediment for restoration of salmonid populations within this region. This will involve reversal of warming trends wherever possible and no additional warming of stream temperatures from anthropogenic activities. All agreed that this is a complex task and no simple solutions will fit all situations.

Some of the questions raised by the Peer Review Panel may result from an incomplete understanding of this draft Criteria. We understand that these ideas are still evolving and the draft documents may not have captured clearly the intent of the working group. The panel hopes that this review will highlight areas of ambiguity and provide constructive suggestions for inclusion into the final Criteria.

The two-stage Criteria is a new approach with which few are familiar or comfortable. For improved understanding by the public, natural resource managers, policy personnel, and reviewers, EPA should explain in detail how it envisions the states' implementation of the Criteria and procedural steps for streams that are not in compliance. The Peer Review Panel inferred that the temperatures in Appendix A could be the region-wide standards, as a default. And that Appendix B provides a preferred technique, but not the only one, to customize standards to specific basins. All panelists agreed that a case study or example to show how this new Criteria would be applied at a site would be helpful to begin to understand the applications and ramifications of a new Criteria. The panel also encourages EPA to include an example of the time line and detailed information on the steps involved in setting a standard for that case study.

Historical conditions (not historic) are an interesting theoretical context. Historical landscape conditions can serve as a foundation for modeling differences between present temperature regimes and those that existed before anthropogenic impacts. But historical conditions have not been developed for most of the region and where they have, there are still many considerations in using these vegetation layers for inputs to models. The only place that the panel knew of having a quantitative spatial database of historical riparian vegetation in the PNW is the Willamette River Basin. The effects of other factors (such as channel morphology, instream wood and sediment) that also influence stream temperature, would need to be incorporated into models. These conditions may be even more difficult to reconstruct than historical vegetation distributions.

The focus of Appendix B on modeling to create a standard for a basin was a concern for several panel members, who were very uncomfortable with the use of models to set standards. They felt that the emphasis should be on a standard that is a simple numerical threshold that represents the potential to support beneficial uses and that modeling is a method to evaluate landscape potential for restoration during implementation phase. Other panelists were not as uncomfortable with the use of models to set standards, but thought that the time and effort required to implement Appendix B may be excessive. All wondered about the time frames required for implementation of the new Criteria and if this methodology would actually delay identification and implementation of restoration strategies on temperature-limited streams.

The proposed Criteria have multiple strengths and weaknesses; the ones considered most important by the panel are those listed in the table below. As is often the case, for every listed strength of the Criteria, there was an opposing weakness. These are presented with the hope that when the Workgroup revises the Criteria, they can present examples of methods to address the weaknesses. The panel discussed these strengths and weaknesses at length and further addressed pros and cons in answer to Question 6.

+ Strengths	**- Weaknesses**
+Thresholds that vary in a spatially explicit manner representative of landscape variability	-Thresholds that vary within a basin will be confusing and expensive to evaluate
+Multiple lines of evidence results in more realistic standards	-Modeling is required as one of the lines of evidence and will delay the time required for states to set standards
+Use of life stage specific temperature standards is sensitive to biological requirements	-More complicated than using only summer maximum temperatures
+Use of historical distribution of species accounts for anthropogenic impacts and reduction in species ranges in present times	-Historical distributions of species are not known for many streams
+Use of historical range of variability in stream temperatures allows adjustment of thresholds to regions or streams with warmer temperature	-Historical ranges are not known for most streams
+Appendix B methodology incorporates uncertainty analyses to examine whether physiological tolerances are within historical range of temperature variability at a site	-Uncertainty can be characterized in multiple ways and used to delay approach or set thresholds that do not protect beneficial uses
+Allowance for human use recognizes that some areas will not be restored	-Allowance for human use is arbitrary and difficult to define

The new Criteria was thought by some panel members to focus more on potential future research and management plans rather than encouraging states to move forward with existing approaches that may meet the intent of the regional guidance. Watershed councils are valuable players in this process, especially because they encourage voluntary implementation of restoration activities, and methods that involve local participation are generally most beneficial to the ultimate goal of restoration of stream temperatures. Panel members suggested that the Criteria could contain specific guidance on how existing modeling efforts by the states can be modified or expanded upon to be in line with the new Criteria. The suggested calculation of range of variability of stream temperatures (called confidence intervals in the Criteria) highlights the importance of bounding model output by uncertainty calculations. The panel thought that not trying to create just one longitudinal temperature trend but viewing longitudinal patterns as a range of variability was a very good idea and could be suggested to states as a methodology to be included in the current modeling efforts.

There was quite a bit of discussion whether it would be possible to remove stream temperature as a limiting factor for beneficial uses with less monitoring, modeling and agency support. Heuristic approaches could potentially be very useful for this. The panel wondered if simple models could be used to prioritize where there might be the greatest thermal response to restoration, in order to more quickly achieve reduction of summer maximum stream temperatures. Could potential 'hot spots' within a basin be located using existing GIS coverages to identify areas where stream temperatures monitoring should be prioritized based on land use practices that are known to contribute to increased stream temperatures? It was suggested that magnitude of diel fluctuations might be an indication of impacted streams, which therefore may have the greatest potential for restoration. A simplified effort might also be able to identify which areas have high quality stream temperature regimes and need protection from further degradation.

An issue not addressed within this guidance, but which has been pointed out that the states need to address, is protection of existing, high-quality, refugia habitat. It is difficult to suggest thresholds or distributions because defining and protecting refugia is dependent upon the scale of the feature as well as the co-occurrence of other necessary habitat features. Spatial scales of refugia range from pockets of cold water to reach scale to subbasins that have cooler stream temperatures. Access to refugia is more critical in degraded habitats and is not an issue in high quality habitats because the definition of physical refugia implies an area of good conditions within an otherwise poor quality matrix. Also not addressed within the Criteria are guidelines for connectivity among high quality habitats and protection of 'core areas'.

The section on allowance for human use was vague and needs to be clarified. Although the Peer Review Panel recognizes that there are features on the landscape that influence stream temperatures but that will not be moved (such as Portland), it is not clear what will happen when human use conflicts with a standard. Blanket allowance for human use could potentially trump all standards; therefore, the section on blanket allowance should be very clearly defined and should be reviewed by a policy peer review group, rather than our technical peer review group.

Throughout the discussion, there was considerable confusion about terminology and what was meant by different terms. In addition, some terms were not used in a consistent manner in the Criteria. Defining the following terms and/or replacing them with other defined terms within the document would help to clarify future discussions:

optimal
confidence intervals, bounds of uncertainty, model error, data error
target, threshold
guidance, criteria
standard
viable populations
climatic -meteorological and hydrological

In summary, the Peer Review Panel suggests the following:

1) Provide an landscape example of this new methodology
2) Clearly lay out the time frames and steps for compliance by states
3) Identify potential alternate strategies that build on existing efforts by states
4) Identify methodology for delisting streams which meet temperature Criteria
5) Set guidelines for determining the number and minimum size of streams within a basin that need to be modeled
6) Clarify the guidelines for allowance for human use
7) Define terms clearly and use them consistently throughout the Criteria

Section 2.

Key questions asked of the Peer Review Panel during review of EPA stream temperature criteria, Nov 8, 2001

Questions concerning Appendix A

1a) *Given that this approach doesn't require optimal temperatures in all streams, all of the time, do you support our effort to focus on identifying the upper optimal temperature for each species and life stage, rather than the upper thermal tolerance?*

1c) *Do you agree with our characterization that these values represent the upper end of the optimal range?*

4) *To address uncertainty associated with physiological thermal preference data, we chose to identify the warmer boundary of optimal thermal conditions. Is this approach reasonable? If not, what method(s) can you recommend for addressing the uncertainty in physiological thermal preferences of fish?*

The setting of more conservative guidance with flexibility in applying that target on the landscape should have a similar 'net effect' to a less conservative guidance that is paired with a more rigid landscape application. Much higher values could have been chosen with the understanding that they may lead to high mortality but not complete extinction in many settings. Alternatively, very low values could have been selected in an attempt to be overly conservative. Instead, the 'upper optimal temperatures' were selected to aim for the highest production (biomass) potential for salmonid populations, i.e., simultaneously maximizing growth (via bioenergetic/metabolic pathways) and minimizing population-level, stress related mortality or migration. It is easy to forget that many growth studies report growth as a function of water temperature for fish that are still alive, failing to report those that died during the course of the experiment. Thus, what we really most often see is growth of individuals, not of populations.

The Peer Review Panel agreed that the suggested methods and optimal values are a reasonable and appropriate approach. Scientific uncertainty about definite thresholds of physiological response, as well as variability in salmonid responses to high stream temperatures as a function of environmental conditions must be acknowledged. Given this uncertainty and the threatened/endangered status of some of the salmonid species, selection of the warmer bounds of 'optimal' provides a conservative target for stream temperatures and supports recovery goals for T&E populations.

However, the use of the word “optimal” resulted in quite a bit of discussion. Its use may be problematic, and could lead to much confusion and debate. Whichever term is selected should have a clearly defined meaning for this set of documents. It is not that the definition of optimal is not workable (optimal was used in the same context in Oregon’s 1993 Triennial Review) but it is a value-laden word where every reader has different connotations of its definition. Alternate terms, such as “adequate” or “appropriate” or “suitable”, could be considered, although these suggested words are also value laden. Elliott (1981) defined “optimal temperature range” as: the range over which feeding occurs and there are no external signs of abnormal behavior, i.e. thermal stress is not obvious. Optimal, as used in these draft criteria, may not be what was defined by Elliott, but also was not clearly defined to be based on ecological responses or performances. Therefore it could be misconstrued to be those temperature conditions where specific indices such as growth, fecundity, or swimming speed are maximized. The panel strongly agrees that the suggested thresholds should be temperatures that protect multiple beneficial uses and not be limited to specific indices or responses. EPA should provide a clear definition of the word that will be consistently used within these documents as well as highlighting the ideas supporting this choice.

Several reviewers mentioned that they thought it would be helpful to have clear protocol of a potential method to incorporate new findings from published studies of fish thermal behavior and preferences. Is there a mechanism that could be defined by which new information can be incorporated to existing temperature thresholds?

1b) *Do you agree with the way we tried to look at acute or short term effects to set daily maximums (MWMT values) and chronic effects to set mean values (MWAT)?*

These types of values have been shown to be highly correlated. The MWMT focuses on highest temperatures a stream attains, while the MWAT contains information about both the maximum and minimum temperatures. The distinction between the effects of chronic versus acute exposure is not clearly defined in the Criteria, nor would the responses necessarily be addressed by using these correlated temperature indices. But the reviewers did not see a problem with having both metrics identified, as long as they are consistently defined and used. It could create a problem if it happened that one metric was exceeded but not the other, in terms of listing a reach. Several panel members suggested that there may be much to learn about the role of daily minimum temperatures for survival in extreme conditions and MWAT does incorporate more than just daily maxima.

Clear definition of all terms would help to eliminate potential confusion. For example, how is the mean calculated? Is there a minimum amount of data required? Is it an average of weekly max and min, daily max and min, or of 4 measurements or 10 or 48/day?

For consistency among guilds, the daily maxima suggested for char should be changed to weekly metrics. There is less scientific certainty regarding fish responses to one hot day as opposed to repeated exposure.

Although these values are 'estimates' of upper temperature preferences, and are not purported to represent rigid (biological?) thresholds, they are being put forward as distinct CWA thresholds (see the subtitle of Appendix A). It was suggested that 'thresholds' be eliminated from the EPA Criteria because the concept has little relevance in natural systems where organisms are responding to a wide variety of environmental conditions, of which temperature is but one. If the temperatures continue to be suggested as thresholds, the third paragraph in Appendix A could be clarified and expanded to incorporate a definition and distinctions, including those in 1a, above.

1d) *What are the considerations in translating laboratory studies of physiological thermal preferences into required thermal regime in the field? For those effects based mostly on lab studies at constant temperatures and high food availability, do these numbers reasonably translate from the laboratory setting to a natural, fluctuating stream?*

No, thermal preferences observed in laboratory conditions do not translate directly to stream conditions, where fish are influenced and responding to many additional abiotic and biotic factors. These considerations were clearly laid out in the technical summaries. Laboratory studies and field "experiments" may have slightly different results for a variety of reasons. For example, age 0+ fish may be observed in warm stream margins largely because they are avoiding predation. In other words, all experiments suffer from their own methodological limitations, so it is important to be clear where temperature response information comes from.

1e) *Is our division into guilds appropriate? Can any guilds be lumped; should any be split? Are our metrics appropriate? Do we need this many metrics or can we simplify?*

Yes, guilds are well defined and appropriate, based on information that is available at the present time. The selected metrics are also appropriate (but see 1b above).

Questions concerning Appendix B

2a) *Is it reasonable to believe that we can accurately assess the confidence boundaries associated with the modeled estimates of historical thermal potential in a specific river basin?*

We can create a distribution with a model, but we cannot "accurately" assess whether it is an estimate of historical thermal potential without actual historical stream temperature data. The modeled distribution is not the same as historical stream temperatures, but is a reference condition, based on a variety of modeling assumptions using the best representation of historical landscape conditions and long-term averages of hydrological and meteorological data to parameterize a model.

The Peer Review Panel had considerable discussion whether these suggested modeling techniques produced confidence boundaries, central tendency, model error or model uncertainty. Although it should be possible to make useful assessments of the confidence with which model results relate to observed conditions, the panel believes that that it is not possible to develop anything analogous to formal confidence intervals using historical conditions. The uncertainty, as described in the Criteria, is an attempt to characterize the bounds of our ability to accurately predict historical conditions. Trying to further describe the "cloud of uncertainty" as a confidence interval simply muddies the objective. If a model is used to create confidence boundaries, these boundaries are linked to the functions assigned within the model and are not the same as uncertainties of temperature within the river. Confidence boundaries are generally used to represent 95% confidence intervals, which, under no terms, could models of historical variability in conditions create.

It would be helpful to have one term consistently used to describe this distribution of modeled stream temperatures along the length of a river. However, there is no widely used or agreed-upon approach in uncertainty analysis, and the field is still evolving. Due to these considerations, there is no specific choice of terminology that stands out as being appropriate. That said, all panelists agreed that confidence limit is a poor choice because it has a statistical connotation different than what is being proposed here. One alternative would be to define a standard approach and terminology, for example "uncertainty bound, defined as the region in which XX percent of the predicted outcomes would lie given explicit characterization of parameter variability, model validity (related to how well the modeled process descriptions capture important drivers of the real system), and input data variability".

2b) *If so, what are the sources of uncertainty (model error) that we should be attempting to quantify and what techniques are available to assess each source of uncertainty?*
2c) *If not, why not? What sources of uncertainty can be assessed? Does the assessable uncertainty generally represent most of the total model uncertainty or only a small portion?*

There are several sources of uncertainty in modeling. One type of uncertainty involves model inputs and use of stream temperatures, meteorological or biophysical variables that are not fully representative of the study site, due to high variability over spatial and temporal extents. Model validity, i.e. how well does the model capture the real system, in part reflects our ability to measure the real system, and in part, reflects the validation process relating these measurements to model outcomes. A second uncertainty involves model parameters, i.e. not fully representing the mechanisms or processes, such as those involved in thermal exchanges and hydrologic transport. There is often high uncertainty concerning how well are these known and how variable are they in the actual ecosystem. Many dynamic and/or spatially explicit models have "hidden" numerical errors resulting from the representation scheme used; these are rarely discussed or quantified, but can be quite important. In general, attempts to reduce model uncertainty by increasing the complexity of the model also increase parameter uncertainty, because more parameters must be estimated.

Model uncertainty can be assessed by modeling existing conditions and comparing the model results to observed data. However, limited data will be available to describe past conditions, so parameter uncertainty will be high, and it will not be possible to reconstruct historical temperature regimes with high accuracy. Many methods for examining the propagation of uncertainty through models exist (e.g. Monte Carlo analysis, Bayesian methods, GLUE (Beven et al.)), and no single method is universally accepted as appropriate for all situations. However, all these and similar methods do in fact provide improved understanding of both the ranges of likely modeled outcomes and the level of confidence with which the modeled results may be believed. This is an evolving field, and application is hampered by our incomplete ability to characterize sources of uncertainty. In any case, a reasonably complete uncertainty analysis, while furthering the interpretability of model outcomes, will require significant effort both in characterizing sources of uncertainty and developing sound methods for propagating uncertainty through the model.

It may be that the only best way of trying to estimate historical "thermal potentials" for a given stream system is with the use of models, but our capability to know the correctness of those estimates are limited by our lack of historical environmental data (temperatures, flows, climatic conditions), our lack of experience in modeling stream networks or whole basins, and our lack of application of such models to basins in good ecological condition (where modeling uncertainties and errors do become readily apparent).

3) *Based on our project goals, we believe a temperature standard must establish how cold water should be, the spatial and temporal pattern of cold water availability, and how much cold water must exist. Is this assumption credible? Is there a more straight forward yet credible means of determining how much cold (optimal) water is necessary and when/where it must exist in order to support viable salmonid populations?*

Protection of beneficial uses is the driving factor. Because salmonids are selected as representing sensitive cold water species, then the temperature of stream water must be not be a limiting factor. The panel discussed whether the criteria are sufficient to achieve salmonid recovery goals. All recognized that numerous factors in addition to stream temperature will influence salmonid recovery. The spatial and temporal extent of stream water at a suitable temperature is an important component and a credible assumption for protection of beneficial uses.

It was not clear in what way 'how much cold water' was different than spatial and temporal patterns? Does it refer to stream discharge? It would seem that concerns over discharge would be covered by setting distributions and thresholds. If water was diverted and flows lowered, then temperatures at critical peak times would be higher than if there were more volume in the stream.

Use of the term "viable salmonid populations" suggests a formal viability analysis, which is not part of this approach. This guidance should not be tied to or rely on NMFS population viability (VSP), which is early in its development and is untested. 'Goals' as outlined previously by EPA should be included in the Criteria to clarify terms. This meeting and previous Peer Review Panel meetings had considerable discussion about linking stream temperature guidance to recovery of salmonid populations. Beneficial uses need to be protected, and high stream temperatures should be removed as a impediment to salmonid recovery. However, evaluation of the success of this temperature guidance should not be dependent upon densities of future salmonid populations.

The panel also discussed and suggests that approaches that do not rely so exclusively on detailed stream temperature models be considered as an additional alternative. While mechanistic models of stream temperature have reached a level of maturity where they do a good job describing the "physics" of energy fluxes contributing to stream temperature dynamics, applicability to the types of analysis proposed here are currently restricted by the availability of spatially explicit datasets to drive the models, lack of understanding and characterization of some processes at a watershed scale (e.g. groundwater and hyporheic flow), and ability to characterize and propagate uncertainty in these models. More qualitative methods, e.g. heuristic methods capturing expert judgment, perhaps in combination with more localized knowledge about particular strains, channel conditions, etc., and more readily available spatial datasets, may be useful in quantifying the spatial distribution of habitat required to protect the beneficial use and consequently to the standards development process, particularly in light of the multicriteria approach proposed here. Such heuristic methods are well developed and can be readily applied in a spatially explicit manner, although they have not been used explicitly for stream temperature and/or habitat analysis. While not necessarily providing the perceived "rigor" of a more mechanistic approach, in fact this perceived rigor from modeling may be misleading. Heuristic approaches have the advantage of more straightforward uncertainty accounting, would be able to effectively capture the types of considerations discussed by the panel in the preceding paragraph, and are more accessible to stakeholders.

5) *In order to determine whether a stream was ever capable of providing thermally suitable habitat, we propose comparing physiological optimal temperatures to confidence boundaries of modeled historical thermal conditions in a median year. Is it appropriate to make this comparison under median conditions? If not, how should the comparison be made?*

As noted above, the panel believes that the data available to describe past conditions will seriously limit the ability to model past conditions with good accuracy and precision. Although modeling may provide information useful for a decision whether a particular stream historically had a thermal regime suitable for salmonids, other kinds of evidence should be considered as well. It may be useful to distinguish geographical areas in which streams are 'presumed' thermally suitable or unsuitable or where no presumption is made for a particular guild, in order to focus efforts in a timely manner on temperature-limited streams for salmonids.
Are they confidence boundaries or zones of uncertainty? (see discussion following question 2 above)

Median may be a confusing term for historical conditions, especially if long-term historical meteorological and hydrological data are not available. It would be more clear to specify that the input would be the conditions, rather than trying to select a year in which there were median conditions. Choosing what conditions to model for water quality standards will continue to be a challenge; to keep it simple but still represent "reality" – which is never simple.

As with ‘optimal’ temperatures, there are trade offs in considering median conditions. ‘Less protective’ thresholds/guidance that always apply might be similar in implementation over time compared to ‘more protective’ thresholds/guidance that only apply some years. To protect resources and beneficial uses, it is reasonable to be concerned with a hot, dry year’s impacts on stream temperature regimes. However, presently, temperature standards do not have to be met in those types of extreme years.

Modeling during extreme conditions, in addition to median, is an important step in validating models and evaluating their sensitivity and potential limitations. If models do not accurately estimate stream temperatures during extreme meteorological or hydrologic conditions, this will highlight gaps in our understanding of mechanisms and/or faulty implementation of processes within the model. Ideally, models would be run for as many years as possible with existing data and continue to be updated with new data. Low variance of temperature regimes among years may be an indicator of resilient streams systems and high variance indicative of sensitive streams.

6) *Is the proposed approach apt to provide substantially less protection (higher risk of extinction) to salmonids than a more common species/life stage approach?*

Appendix A is a species/life stage approach. That the thresholds suggested in Appendix A can be modified using multiple lines of evidence is a major part of this approach. As well, this approach incorporates longitudinal dynamics where headwaters might be colder than thresholds and downstream reaches might be warmer. But whether that could be argued to constitute substantially less protection and higher risk of extinction depends on definitions beyond the scope of this review.

There are pros and cons to any approach. This approach is flexible because it does not apply one threshold everywhere, but examines landscape, historical and regional dynamics. This approach also provides potential protection and non-degradation of existing cold water headwater habitats, even if they are colder than the thresholds. Preserving this variability within the landscapes is a strong plus of this methodology.
Cons -The methodology in Appendix B may be far more cumbersome, time consuming, and expensive than an Appendix A-type standard. However, it allows more "wiggle room", which

may be good or bad. The uncertainty involved in modeling historical range of variability could be taken advantage of and standards established that are not protective of beneficial uses. In addition, this approach does not establish guidance for protecting cold water refugia along main stems or maintaining connectivity or minimum distances between cold water patches, which have been noted by EPA to be deficiencies of present standards.

7) *Developing a reliable model of thermal potential and associated uncertainty would require a 'good" model. Are there a set of performance guidelines that would help ensure that a selected model is state-of -the-art and reliable? What is the best means to validate any model?*

Many factors are critical for "good" models. They need to be based on accepted mechanistic techniques with explicit identification of processes considered, explicit statement of assumptions, well documented model processes, and code available for examination. The appropriate spatial and temporal scales should also be stated. Ideally, good models are "supported" by having an identified party to handle technical assistance, bug fixing, training, etc. It is desirable if the models themselves can calculate goodness-of-fit metrics so that they can be compared across applications in a standard way.

In addition, the models need to be able to match the real world within some a priori stated degree of accuracy. 'How closely?' is a difficult question, and relates to primarily to one's confidence that the model can be used for its stated purposes (for example, identifying cost-efficient solutions, back casting historic or forecasting future mitigation conditions, etc.). Each "successful" application, particularly in novel circumstances, adds to confidence that the model can be applied elsewhere with high fidelity and trust.

There is a "classic" validation procedure that involves calibrating the model with one set of data and "validating" the model with an "independent" data set. Almost never does the model behave quite as well with the validation data set. Modelers often then, rather arbitrarily, pool the data, recalibrate and use the model, or discard the model as untrustworthy in this case. Not that the panel is recommending this type of validation for each and every study area where models are used. Generally, if the model has been shown to be effective in similar circumstances and landscapes, a strict validation may be unwarranted.

Finally, there was discussion about "over calibrating" a model to specific data, only to find that it doesn't seem to work so well under altered conditions. It is appropriate to recognize and document the performance limitations of any model. A variety of statistics (probable error, r-square, mean absolute deviation, etc.) can be used, but all have their limitations and none should be used alone. Also discussed was that validation is not a one-time effort. As additional data is obtained through monitoring, re-examination of a selected model and the fit of output to observed values would lend insights to model processes and their accuracy in driving the results. And just because the model was improved, it does not mean that what was learned in a prior application was wrong. Reckhow et al. 1990 (Statistical evaluation of mechanistic water-quality models. J. of Environmental Engineering, 116:250-268) was recommended for more on this topic.

8) *If a temperature model is used to estimate thermal potential in a subbasin, is it necessary to apply the model to all the streams within the subbasin? That is, can a reliable estimate of thermal potential be derived by focusing the modeling on the mainstem? Is there a way to integrate the potential of the tributaries in the overall mainstem thermal profile?*

The amount of work required to model and monitor all streams within a subbasin would be overwhelming. However, modeling and monitoring the mainstem alone may not provide protection of important habitat and potential refugia for salmonids, our indicator of beneficial use. Therefore modeling and characterization of some tributaries is necessary. These could be selected through stratified sampling methods using geology, land use, volumetric inputs to the main channel as well as other criteria. Initial temperature monitoring and examination of fish distributions may highlight the extent to which specific tributaries and mainstem reaches are functionally linked and providing critical resources or habitat for salmonids. Without some information on the tributaries, it would not be possible to integrate their contributions to the mainstem thermal potential. Guidelines in the Criteria would help clarify the necessary spatial and temporal extent of modeling of thermal potential.