Appendix C

Peer Review of the Tahoe Regional Planning Agency's 2015 Threshold Evaluation Report

Executive Summary

September 14, 2016

Submitted to:

The Tahoe Regional Planning Agency

By:

Sonia A. Hall, PhD – Associate Scientist

Christine M. Albano, MSc – Lead Scientist

and

Brett G. Dickson, PhD – Chief Scientist

11050 PIONEER TRAIL • SUITE 202 • TRUCKEE, CALIFORNIA • 96161 • (p) 530.214.8905 | (f) 530.214.8907
5 OLD TOWN SQUARE • SUITE 205 • FORT COLLINS, COLORADO • 80524 • (p) 970.484.2898 | (f) 970.484.1069
info@csp-inc.org | www.csp-inc.org

Conservation Science Partners coordinated an independent review of the Tahoe Regional Planning Agency's (TRPA) 2015 Threshold Evaluation Report between April and August 2016. Twelve experts from ten different universities, one federal agency, and a private consulting firm—two of whom had already participated in the 2011 Report's peer-review process—completed the review of the Introduction, Methodology, and nine topic chapters: Air Quality, Water Quality, Soil Conservation, Vegetation Preservation, Fisheries, Wildlife, Scenic Resources, Noise, and Recreation.

In addition to a brief description of the peer-review process, this document provides a synthesis of reviewers' comments, focusing on:

1. Their assessment of the monitoring program for each threshold category.
2. Recommendations on how to improve the monitoring program for the current threshold standards.
3. Recommendations on how to improve the threshold standards themselves.

Reviewers commended the agency not only for investing in status and trends monitoring to inform their programs and decisions, but also for reflecting on the confidence that each measure provided.

Overall, the reviewers' assessment of the 2015 Threshold Evaluation Report chapters varied, mostly depending on what chapters each one reviewed. With a few exceptions (where reviewers critiqued a chapter from different perspectives), reviewers of the same chapter generally agreed on whether the monitoring results supported the conclusions drawn on the status and trends of that particular environmental resource. The assessments ranged from reviewers stating that conclusions were adequately supported by the data, such as for the Air Quality and Water Quality threshold categories, to reviewers voicing concerns that the current information cannot support management decisions, such as for the Recreation threshold category.

Reviewers provided recommendations for improvement that highlighted eight themes, including the need to:

1. Define a clear and complete sampling design for each threshold standard.
2. Incorporate advances in technology, knowledge, and methods.
3. Identify indicator-specific analytical methods for evaluating trends.
4. Distinguish between insufficient data and no trend.
5. Improve how confidence in results is determined.
6. Leverage data to better inform Basin-wide decisions and need for action.
7. Improve the link between standards, monitoring results, and actions.
8. Clarify definitions and terminology.

In addition, reviewers provided recommendations for improvement of the standards themselves that highlighted three themes, namely the need to:

1. Articulate explicit objectives for each threshold category.
2. Use advances in science, knowledge, and methods to review and revise threshold standards.
3. Monitor factors that act as drivers of change in the Lake Tahoe Basin.

Reviewers generally identified greater deficiencies in the threshold standards themselves than in the processes used to evaluate them. Recommendations to improve the monitoring program for current standards would be easier to implement once TRPA completes its initiative to evaluate and modify their threshold standards.

The rest of this document provides a brief description of the peer-review process, summaries of findings for each chapter, more detailed descriptions of reviewers' recommendations for improvements to the monitoring program and to the threshold standards themselves, and a conclusion with recommendations to help strengthen and potentially streamline TRPA's monitoring program, so results can more effectively and efficiently informing TRPA's actions.

The Peer Review Process

The purpose of the 2015 Threshold Evaluation Report peer-review process was to determine whether appropriate scientific methods and the best available science were adequately applied to evaluate the status and trends established in Resolution 82-11, so as to meet the reporting requirements described in the Code of Ordinances Chapter 16 (see Peer Review Charge). Conservation Science Partners (CSP) identified and recruited two expert reviewers for each topic chapter, compiled the reviews once received, submitted them to the Tahoe Regional Planning Agency (TRPA) staff, and authored this synthesis of key findings.

The peer-review process mirrored the process scientific journal editors use to obtain peer review for manuscripts they receive. In our case, we identified experts on the particular topic and invited them to review specific chapters. Our selection of potential reviewers was based on:

(a) Relevant topic expertise,
(b) Scientific standing, reflected in their publication record and recommendations from other experts,
(c) Experience and interest in monitoring, and how it informs planning and policy decisions, and
(d) Independence from contributors to the 2015 Report itself (see Contributors List).

We also considered whether (e) reviewers had expertise relevant to more than one threshold category, and whether (f) experts had previously provided TRPA with high-quality reviews. These last two considerations allowed us to more efficiently complete the review of the whole Report, and provide some continuity in perspective from the last peer review, in 2012. If invitees were unable to complete the review in the time requested, we asked that they recommend other experts in their field.

Each of the twelve reviewers critiqued one or two topic chapters, as well as the Introduction and the Methodology. TRPA did not include the implementation, recommendations, and conclusions chapters in this review, a difference from the peer review of the 2011 Threshold Evaluation Report. Each topic chapter received at least two expert reviews. In addition, Dr. Sonia Hall, associate scientist with CSP, provided comments on the majority of the chapters from a generalist's perspective.

The panel of reviewers (Table 1) was charged with evaluating and commenting on the information, analyses, results and recommendations contained in the 2015 Threshold Evaluation Report (see Peer Review Charge). They were also asked whether the Report contained the information necessary to make informed environmental decisions. Each reviewer provided a written review, which generally consisted of an overall assessment of the chapter, followed in many cases by specific comments about particular sections (see Individual Reviews). Once we received all the reviews for each chapter, we submitted them to TRPA staff. Only at this point did we share with TRPA the names of the reviewers. A few of the reviewers also offered additional resources, including publications or alternative analyses, which we also forwarded to TRPA.

Table 1. Peer Review Panel for the 2015 Threshold Evaluation Report

Reviewer	Affiliation	Topic chapters reviewed
Jesse Barber, PhD	Boise State University	Wildlife, Noise
Dave Beauchamp, PhD	USGS Western Fisheries Research Center	Water Quality, Fisheries
Derek Booth, PhD	Univ. California, Santa Barbara	Soil Conservation
Robert Burns, PhD	West Virginia University	Scenic Resources, Recreation
Stephanie Guildford, PhD	University of Minnesota, Duluth	Water Quality
Gary Hunt, QEP	TRC Environmental Corps	Air Quality, Noise
Jeffrey Marion, PhD	Virginia Tech University	Scenic Resources, Recreation
Robert Naiman, PhD	University of Washington	Vegetation Preservation, Fisheries
Barry Noon, PhD	Colorado State University	Wildlife
Kevin Rose, PhD	Rensselaer Polytechnic Institute	Water Quality
Scott Spak, PhD	University of Iowa	Air Quality
Carol Wessman	University of Colorado, Boulder	Soil Conservation, Vegetation Preservation

Assessment of the 2015 Threshold Evaluation Report's Monitoring Program

The reviewers' assessment of the 2015 Threshold Evaluation Report chapters varied, mostly depending on what chapters each expert reviewed. **With a few exceptions, reviewers of the same chapter generally agreed on whether the monitoring results supported the conclusions drawn on the status and trends of that particular environmental resource. The exceptions were generally due to reviewers critiquing the chapter from different perspectives.** The specific summaries for each chapter are driven by the expert reviewers' comments[1].

***INTRODUCTION*:** All twelve reviewers reviewed the introduction chapter. **Reviewers generally found the introduction to be adequate and sufficient.** A couple of reviewers commented on the lack of support for statements, though their suggestions generally focused on the value of providing links to supporting data and information, rather than considering this a critical flaw. Another pair of reviewers praised the use of the reporting icons. Reviewers provided a number of suggestions for improving the introduction:

(a) Provide additional "big picture" context, including population and socioeconomic information, details on the drivers of environmental trends, and the environmental challenges TRPA is facing; and
(b) Define key terminology used in the report, including "threshold", "parameter", and "targets", among others.

***METHODOLOGY*: The majority of reviewers considered the methodology chapter described adequate approaches for determining the status and trends of indicators relative to the adopted standards, though most reviewers also had specific questions and recommendations**, targeting:

(a) The use of linear regression as the main method for evaluating trend,
(b) The use of r^2 and p-values to determine confidence in said trend, and
(c) The temporal scale used in determining trends.

Three reviewers, however, were not satisfied with the overall methods. Many of the issues they raised focused on a need to better understand what the indicators were meant to inform, what the ecological rationale for the threshold standards was, and how these questions should drive methodological decisions, rather than having a standard approach across the board. All three of these reviewers also raised issues with how confidence was addressed, though they pointed out the value of including measures of confidence. This is an improvement compared to the 2011 Report, as a reviewer who participated in that earlier review pointed out, given that confidence in this 2015 Report was not biased high when insufficient data were available or standards were lacking.

***AIR QUALITY*: Both reviewers agreed that conclusions reached were supported by available data, with a few exceptions.** Both reviewers were also generally supportive of the recommendations for modifications provided by the chapter's authors. Where authors described options being considered, both reviewers suggested preferred alternatives. In addition, reviewers provided a number of recommendations, including improved collection of emissions and nitrogen deposition data, development of new odor standards, the continued pursuit of numerical standards, additional methods (such as modeling and remote sensing) and modifications for the visibility threshold standards, and the need for better ties to updates in regulations and in scientific analyses carried out by others.

***WATER QUALITY*: All three reviewers generally agreed that the conclusions reached were supported by available data, with few exceptions.** All three also explicitly supported some of the recommendations for modifications provided by the chapter's authors. The recommendations from these three reviewers were

[1] Dr. Hall provided additional comments from a generalist's perspective. Her recommendations are synthesized in the chapter summaries and the recommendations for improvements. They generally aligned with the perspective of other reviewers.

[2] In the reviewer's words (in reviewing this synthesis document): "Having a well-defined framework for the major evaluation interval for

generally well aligned, including calls for better descriptions and more clarity on terms and methods, and improved discussion of the ecological understanding underlying statements. Some examples include the need for better understanding of which nutrients are most limiting; the importance of linking indicators to understand the relationship between runoff nutrients and lake response; and a call for analyses to understand the relative contributions of different factors impacting particular indicators.

All three reviewers pointed out the need for targeting monitoring at a temporal scale that reflects the dynamics of these factors, particularly looking at seasonal changes, changes related to flow events, and the importance of episodic or extreme events in driving dynamics. One reviewer also voiced concern about the need to use better methods to evaluate trends, and distinguish lack of trend from insufficient evidence to confirm a trend. The main concern on the threshold standards themselves, voiced by two reviewers, was whether these threshold standards would be sufficient to improve water quality under a changing climate.

***SOIL CONSERVATION*:** The two reviews had notably different focus. **One reviewer mostly disagreed with the conclusions drawn by the authors**, though his concerns were mainly related to the nature of the threshold standards themselves (including their basis in 30+ year old understanding of sedimentation processes), and therefore their inability to support conclusions about the status of soil resources in the Basin. **The second reviewer focused more specifically on whether the conclusions related to the current threshold standards were supported, and generally found that they were**, and that she agreed with authors' recommendations for addressing the factors currently limiting those evaluations. Recommendations for improvement of the monitoring program included a process for updating critical benchmarks for the current threshold standards, as well as the need for a better benchmark, based on best available science, for the Stream Environmental Zones.

***VEGETATION PRESERVATION*: One reviewer found that the conclusions related to the current threshold standards were generally supported. The second reviewer concluded that for many indicators the temporal coverage and validity of the data were insufficient**, which is reflected in the lack of trend determinations for a number of the standards. Both reviewers voiced concern about the major evaluation interval. In reviewing this synthesis, one reviewer emphasized that the determination of evaluation intervals for each standard should be driven by the most ecologically meaningful sampling frequency for that standard[2].

Both reviewers generally supported the authors' recommendations for addressing the factors currently limiting the evaluations, with the exception of those relating to the major evaluation interval, described above. They also made their own recommendations for improvements to the current monitoring program, including using aerial photos to address the limited temporal coverage, having descriptions of lessons learned so far, and descriptions of the plant communities that were not included (such as phytoplankton, given the emphasis on lake clarity).

Both reviewers had recommendations focused on the threshold standards themselves, including the need to take a landscape perspective by including landscape metrics and a more refined understanding of types of disturbance, and the need to consider climate change impacts, as well as developing quantitative objectives, providing better ecological rationales for threshold standards, and giving more attention to critical components of vegetation, such as the understory, non-native species, and riparian forests.

***FISHERIES*:** One reviewer stated that conclusions relating to the limited, coarse habitats monitored were generally supported (with some concerns about trends in littoral habitat). However, he concluded that the chapter overall lacked clarity (e.g. quantitative definitions) in the evaluation categories, the monitoring was in many cases limited to non-existent (e.g., in-stream flow), and that the limited amount of restoration that has been attempted was disappointing. In addition, **both reviewers highlighted the current threshold standards' inability to inform decision-makers on the status and trends in fish populations**, and generally supported

[2] In the reviewer's words (in reviewing this synthesis document): "Having a well-defined framework for the major evaluation interval for each standard is important for determinations of trends and for the establishment of monitoring practices."

authors' recommendations for improvement. Most of their comments focused on needing to expand the threshold standards themselves to other factors, including fish (relative abundance, distribution, size structure), pelagic habitats, non-native species, stream temperature, and instream flow requirements for fish. Both reviewers also highlighted the need to improve the understanding of the ecological rationale behind the standards.

***WILDLIFE*: One reviewer found that wildlife was generally well addressed in the 2015 Threshold Evaluation, though statistical approaches to quantify wildlife populations were not his specific area of expertise.** He provided three key recommendations for improvement of the threshold standards: establish standards for songbirds and bats, as they are relatively easy to monitor and can reveal broader patterns of habitat quality; include standards that monitor impacts of roads on wildlife; and consider noise and light in definitions of disturbance free zones. **The second reviewer concluded that the sampling design was deficient, leading to an inability to draw conclusions about the status of selected wildlife, and that the analytical approach was not the most appropriate to determine trends.** He also recommended a surrogate species approach to provide broader information on all wildlife species. In addition, he provided an in-depth explanation to help improve the sampling design, as well as a description of alternative methods for analyzing wildlife population trends (including an example reanalyzing data in the 2015 Report).

***SCENIC RESOURCES*: Both reviewers considered that the indicators were comprehensive and the logic for evaluation was appropriate.** One reviewer recommended using stakeholder perception data to determine if users (those who enjoy the scenic resources) share the conclusions described in this Report. The second reviewer's recommendations focused on illustrating what are inherently subjective standards of scenic quality, suggesting the Report provide more examples of the challenges and potential solutions, illustrated with before-and-after photographs.

***NOISE*:** The two reviews of the noise chapter focused on two different perspectives: impacts of noise on people, and impacts of noise on wildlife. **The reviewer who had reviewed this chapter for the 2011 Report found big improvements in the monitoring program since then, though the data are as yet insufficient to support statements on program effectiveness.** He continues to recommend changes to the threshold standards, stating they are too many, too complex, and too resource intensive. He further suggests using existing data to reevaluate standards that may not be attainable, and to focus on annual rather than extreme values to evaluate trends. The second reviewer's biggest concerns focused on the fact that the noise threshold standards are not designed to address wildlife impacts. His recommendations therefore focus on improving the thresholds and indicators to address this limitation. He also requested more details on the monitoring methods.

***RECREATION*: Both reviewers voiced concerns that the current information cannot support management decisions.** Methods were not well described so could not be evaluated, and data for at least two time periods—collected using the same surveys and protocols—is needed to inform trends. Both reviewers support the authors' recommendations to improve the use of the PAOT data, and to use the US Forest Service's Visitor Use Management framework. One reviewer additionally recommended using the Recreation Opportunity Spectrum system for planning and management purposes. A particular issue highlighted was the need for better threshold standards based on a better-articulated guiding statement, which should also include the protection of natural, cultural, and historical resources.

To summarize, all reviewers identified areas where this 2015 Report can be improved. Detailed comments and recommendations are provided in full in each individual review. However, that is only one part of the story. In addition, many reviewers highlighted the importance and value of TRPA's status and trends monitoring effort. Though the peer-review process is inherently critical, reviewers highlighted that TRPA's Threshold Evaluation for the Lake Tahoe Basin resources is in and of itself a significant and valuable investment that few agencies undertake. As one reviewer put it, *"The agency is commended for this clear and extensive approach to indicators of local environmental change, at the leading edge of a regional planning indicator set relative to best practices."*

Improvements to the Monitoring Program for the Current Threshold Standards

Reviewers provided recommendations for improvement that highlighted eight themes across different threshold categories. These are:

1. **Define a clear and complete sampling design for each threshold standard**, to make statistical inferences from the monitored sample. As one reviewer described in detail for the wildlife standards, this includes defining the geographic scope to which inferences need to apply, the sampling units where the relevant indicators could be measured, and the selection of an adequate sample, in time and space, that will actually be measured.

   Temporal scale: Articulating a clear rationale for the temporal scale and timing of when indicators are measured will strengthen TRPA's ability to use the monitoring results to infer the status and trends of the threshold category more broadly, and therefore to inform what actions are needed to improve their status. These timing decisions should also consider whether to use annual averages or extremes. As explained in relation to the water quality chapter, the use of annual averages to analyze trends is problematic when variables are driven by a few episodic events. In addition, the rate of change that can be expected and the horizon for achieving attainment will affect decisions on the temporal scale at which to analyze trends.

   Spatial scale: Reviewer comments highlighted the need to ensure that monitoring is designed at the right scale and extent. For example, the air quality chapter review discussed the importance of understanding nitrogen deposition, a process driven (at least in part) by emissions outside of the Lake Tahoe Basin. And the review of the vegetation chapter articulated the need for landscape-scale metrics to capture trends in processes operating at that scale, such as fire.

   Though the same design may not be appropriate for the different standards in a threshold category, a consistent design can both improve monitoring efficiency, as well as support better ties across standards, and therefore improved understanding of the system as a whole (see recommendation #6, below).

2. **Incorporate advances in technology, methods, and knowledge**. Reviewers highlighted remote sensing technologies, automated sampling of water quality, and improved understanding of sources of sediments as advances that, if used, would improve TRPA's monitoring program.

3. **Identify indicator-specific analytical methods for evaluating trends**, rather than using simple linear regression across the board. Reviewers provided different reasons for needing such indicator-specific methods, including violations to key assumptions of linear regression, and auto-correlation among data. Alternative statistics for confidence should be explored if the data are non-Gaussian. A wildlife reviewer recommended the use of exponential growth models to evaluate trends in wildlife populations. The water quality reviewers recommended the use of Sen's slope for time-series data. Reviewers also recommended the use of structural equation modeling to explain the complex relations between variables, and principal components analysis to determine complex trends and causalities.

4. **Distinguish between insufficient data and no trend.** Some reviewers were concerned about the interpretation given to low r^2 and high p-values obtained from analyzing trends. Though this can be an indication of insufficient data to distinguish a trend given the data's variability, it is also possible that the true signal is that there is no trend; i.e. the indicator is not changing through time. The response to these two reasons for a low r^2 and high p-values would be different. In the first case, better data are needed to evaluate the real trend. In the second case, action is needed to ensure a positive (or negative) trend towards attainment, unless the lack of trend means the standard remains in attainment. It is therefore critical to distinguish between these two reasons. Multiple reviewers recommended consulting with statisticians on the best approaches for doing so.

It is important to note that a reviewer who also reviewed the 2011 Threshold Evaluation Report pointed out the improvement in this Report in the treatment of indicators with insufficient data. In this 2015 Report it is clearer that trends are evaluated, generally, only when enough data of sufficient quality are available. This improvement likely contributed to highlighting the need to now take this a step further, to determine when there really is no trend in the data.

On a related topic, multiple reviewers also highlighted the value of ensuring that, when monitoring methods change, the old and new methods are calibrated. Such calibration supports continued inferences on trends, even as methods improve the way indicators are tracked. One example in the current Report that does that was called out: the cross-calibration of 5.5 years nitrite + nitrate data using old and new methods.

5. **Improve how confidence in results is determined.** Different reviewers identified different aspects of confidence that need to improve. One reviewer took issue with the 2-out-of-3-is-good approach, stating that the poor third one can sometimes trump the other two. He recommended evaluating the weight of evidence using critical professional judgment in addition to (or in replacement of) the current approach. A second reviewer also recommended addressing sources of uncertainty in the context of confidence levels. A third argued that the discussion of confidence needs to be more strongly grounded in statistical theory. It is noteworthy, however, that multiple reviewers pointed out how important it is that level of confidence is an integral part of the evaluation, and commended TRPA for including it.

6. **Leverage data to better inform Basin-wide decisions and need for action.** A number of reviewers made suggestions about how to more effectively leverage the data collected to draw broader conclusions. We realize that the authors of the Report may already be covering many of these in concluding chapters that were not part of this peer-review process. However, we share some of these examples in the hopes that they are helpful in guiding such concluding work, as well as guiding improvements in monitoring in the future.

   (c) *Include an Executive Summary for the 2015 Threshold Evaluation Report* as a whole, to paint a comprehensive picture of the status of the Basin, based on the cumulative evidence from all the threshold standards.
   (d) *Improve the consistency in indicators across indicator categories*. Such consistency would help TRPA understand, for example, how changes in the watershed impact pelagic, littoral, tributary, and stormwater runoff.
   (e) *More intentionally use noise, vehicle traffic, and emissions data to inform wildlife status*, by helping to form hypotheses on needed actions. Noise, light, and roads can impact wildlife, whose response to these disturbances will be different to people's[3].
   (f) *Update each Threshold Evaluation Report by reviewing key findings from peer-reviewed and regulatory studies* published since the last Report.
   (g) *Use a surrogate species approach* to draw inferences about the status of a broader array of wildlife species.

7. **Improve the link between standards, monitoring results, and actions** by explicitly using an adaptive management framework. Reviewers clearly recognized the investment the TRPA makes in monitoring and reporting status and trends. The identification of areas—and even full threshold categories—in need of improvement suggests that prioritizing work and increasing efficiencies would improve the impact of these monitoring efforts. Different reviewers pointed to aspects that TRPA should consider, that would help focus

[3] In reviewing this synthesis, a reviewer of the fisheries chapter also highlighted that recent research is finding quantifiable impacts of light on nearshore and pelagic regions of lakes. This was not included in his original review, as nothing in that chapter triggered concerns about light pollution.

and prioritize monitoring resources in pursuit of determining whether environmental standards are reached. These include:

(a) *Determine when standards are considered unattainable, and articulate what monitoring changes or action this dictates*. One reviewer questioned the current criterion of zero exceedances for single noise events, and another reviewer supported the authors' recommendation to modify the Galena Creek rockcress target because it was considered unattainable.
(b) *Consider what monitoring changes or actions are dictated by standards that have been in attainment for multiple reporting cycles*. Reviewers identified such consistent attainment for the carbon monoxide indicator, as well as recommending changes to wildlife standards, given the consistent results on raptors.
(c) *Investigate the causal links between human and environmental drivers and the status of indicators*. Multiple reviewers recommended the use of supplemental data, published analyses, modeling, or additional analyses to articulate—and where possible quantify—those links. Examples the reviewers specified included the relationship between total phosphorous and turbidity; the use of supplemental data where monitoring is scarce; the use of modeling to simulate values at higher frequency; the relationship between groundwater extractions and instream flows; and the driving sources of sediment.
(d) *Articulate how the results of the monitoring support particular actions*. Reviewers provided multiple examples, such as questioning if there is evidence that habitat is limiting to fish; requesting a distinction be made between ecological and decision thresholds; and proposing an updated framework that includes the development of alternative management approaches to achieve stated goals.

8. **Clarify definitions and terminology.** Semantics can be important, as different uses of certain terms can lead to confusion between those in biological fields and decision-makers. Many of the reviewers pointed out terms that were confusing to them. In a number of cases, these comments arose in response to the authors pointing out the issues they faced in interpreting those terms (such as "disturbed, developed, or subdivided"). Many others, however, focus on key terms used by TRPA, that arise from their initial mandate to develop and track threshold standards, but that were confusing to the reviewers. Examples include threshold, threshold standards, threshold carrying capacity, threshold indicator reporting categories, and targets. We would recommend the use of commonly accepted terminology wherever possible.

Improvements That Would Require Changes to the Threshold Standards Themselves

In addition to charging reviewers with commenting on data and analyses for each topic chapter, reviewers were asked whether they considered the Report contained the information needed to make informed environmental management decisions. Many reviewers responded to this invitation to comment on the threshold standards themselves. In some cases, the threshold standards were the focus of the review (e.g. reviews on soil conservation, fisheries, and recreation). This emphasis, coupled with the knowledge that TRPA's Governing Board has identified the review and update of the threshold standard system as a strategic initiative for the agency to engage in between 2015 and 2020, led us to synthesize reviewers' findings on the threshold standards themselves. We hope that TRPA will find this synthesis useful, as they decide on how to assess, evaluate, and update the threshold standards in preparation for future Threshold Evaluation Reports.

Reviewers provided recommendations for improvement of the standards themselves that highlight three themes across different threshold categories. These are:

1. **Articulate explicit objectives for each threshold category.** Reviewers called for explicit, measurable objectives and a clear chain of logic that explains the rationale for the selected standards, and how the

standards allow TRPA to determine the effectiveness of their programs, and consequent improvement in status. Different reviewers articulated this from different perspectives:

(a) Some reviewers questioned the ecological validity of thresholds, and the need to use best available science to establish those threshold standards. One reviewer praised the rationale for Tahoe draba as an example to follow.
(b) One reviewer stated that quantitative objectives were needed for vegetation. He also pointed out that the benefits of existing flow regimes for fish and other community components had not been demonstrated.
(c) Both fisheries reviewers stressed the potential for current standards to be misleading in terms of condition of fish populations, given that fish are not directly monitored. One of them recommended the inclusion of additional threshold standards that directly focus on the status and trends of littoral fishes and pelagic habitats.

The lack of clearly articulated objectives and rationale for particular indicators is at the heart of the limitations that reviewers highlighted for the wildlife, fisheries, and soil conservation chapters.

2. **Use advances in science, knowledge, and methods to review and revise threshold standards.** This builds on the previous recommendation, highlighting specifics that would need to be quantified, and approaches to help TRPA do so. Reviewers:

(a) Stressed the need to explain the ecological rationale for the thresholds used. These thresholds should reflect a meaningful ecological threshold, one that TRPA is confident would avoid an undesirable change of state of the environmental resource in question. This is a key place to apply best available science.
(b) Encouraged TRPA to continue to move towards fully numerical standards, supported recommendations from chapter authors on such movement towards numerical standards, and questioned the value of management and policy standards as currently presented.
(c) Called for fully articulating how each threshold standard informs action in an adaptive management framework, streamlining standards that may be too many or too complex, and designing sampling to adequately answer the necessary questions on status and trends. These comments point towards a need to prioritize monitoring efforts, given what are expected to be limited resources.

Reviewers also called out specific areas where science has led to important advances in our understanding since the threshold standards were developed in 1982. These include:

(a) Noise and light as sources of disturbance, with potential impacts on wildlife, and on aquatic invertebrates and native fishes.
(b) Sources of soil erosion that impact water quality.
(c) Ecological importance of understory vegetation in forests.
(d) Connectivity and its importance to vegetation, wildlife, and fish.
(e) Temperature effects on fish.

Such advances should be considered as the threshold standards are assessed and revised. In addition, given TRPA's role as a monitoring partner and funding entity, TRPA could also encourage the publication in peer-reviewed journals of relevant findings from both the monitoring and any supplemental analyses carried out to understand causal relationships. Such publications would complement peer-review efforts such as the one summarized in this document, and strengthen the ecological and scientific basis for TRPA's threshold standards.

3. **Monitor factors that act as drivers of change in the Lake Tahoe Basin**, and could lead to direct or indirect impacts on the environmental resources of interest. Drivers of change called out by reviewers include:

(a) Invasive species, with potential consequences for vegetation, fisheries, and air quality.

(b) Microplastics and organic pollutants, with potential consequences for air quality.
(c) Local climate and projected climate change, with potential consequences for species, terrestrial systems, and aquatic systems.
(d) Fire, with consequences for vegetation, water quality, air quality, scenic resources, wildlife, and recreation.
(e) Landscape fragmentation, with consequences for vegetation, wildlife, and likely other categories.
(f) Demographic changes that drive changes to the region's economy, with potential consequences to all threshold categories.

As previously mentioned, we recognize that TRPA is embarking on an assessment of the threshold standards, with the intent to evaluate and modify them. We strongly support this course of action, and emphasize that many of the improvements to the monitoring program that are synthesized in the **Improvements to the Monitoring Program** section will be easier to implement once the recommendations on improving the threshold standards themselves have been carried out.

Conclusion

The peer review of the Tahoe Regional Planning Agency's 2015 Threshold Evaluation Report has provided an in-depth critique and a wide array of recommendations for improvement of the monitoring program for current threshold standards, as well as improvements to the threshold standards themselves. Though reviewers' assessment of the current monitoring program varied across threshold categories, reviewers generally identified greater deficiencies in the threshold standards themselves than in the processes used to evaluate them. Recommendations to improve the monitoring program for current standards would be easier to implement once TRPA completes its initiative to evaluate and modify their threshold standards. As TRPA makes improvements, we strongly encourage the agency to:

1. Support publication of the science and monitoring the agency funds,
2. Pursue independent peer review of the modified and updated threshold standards once the assessment is complete, and
3. Work closely with statisticians and monitoring design experts to improve the monitoring program design for the modified standards.

These actions will help strengthen and potentially streamline TRPA's monitoring program, so results can more effectively and efficiently informing TRPA's actions to preserve, restore, and enhance the unique natural and human environment of the Lake Tahoe Region, while improving local communities, and people's interactions with your irreplaceable environment.

Peer Review of the Tahoe Regional Planning Agency's 2015 Threshold Evaluation Report

Individual Reviews

September 23, 2016

Submitted to:

The Tahoe Regional Planning Agency

By:

Sonia A. Hall, PhD – Associate Scientist

Christine M. Albano, MSc – Lead Scientist

and

Brett G. Dickson, PhD – Chief Scientist

11050 PIONEER TRAIL • SUITE 202 • TRUCKEE, CALIFORNIA • 96161 • (p) 530.214.8905 | (f) 530.214.8907
5 OLD TOWN SQUARE • SUITE 205 • FORT COLLINS, COLORADO • 80524 • (p) 970.484.2898 | (f) 970.484.1069
info@csp-inc.org | www.csp-inc.org

Table of Contents

Note: In addition to their individual review, some reviewers also provided detailed comments in an annotated version of the chapter or chapters they reviewed. Others provided additional materials and resources that they considered useful. These annotated chapters and additional materials and resources were submitted to the Tahoe Regional Planning Agency. For simplicity and consistency, however, they are not included in this compilation of individual reviews.

Biosketches

Peer Review Coordinators

Sonia A. Hall, PhD – Peer Review Chair

Sonia is an Associate Scientist at Conservation Science Partners, and principal of SAH Ecologia LLC. She brings 11 years experience as an applied conservation scientist, with an emphasis on developing science products that directly address stakeholders' information needs. She has experience in partnership building, facilitation and project management, and outreach for science projects. She has coordinated landscape-scale modeling, conservation planning, and ecoregional prioritization projects, including the technical peer-review of the Columbia Plateau connectivity analysis. She was an applied scientist at The Nature Conservancy for 9 years, and has partnered with scientists at universities, state and federal agencies, and other NGOs. She is lead or co-author on 15 peer-reviewed publications, as well as multiple technical reports. She has also reviewed articles for multiple journals, including Frontiers in Ecology and the Environment and Restoration Ecology.

Christine Albano, MSc

Christine is a Lead Scientist at Conservation Science Partners and brings 15 years of research and project management experience. She has worked as a field ecologist, university researcher, and manager of a conservation science and land stewardship program for a regional nonprofit. She has 12 peer-reviewed publications in the fields of Ecology and Conservation Biology, has authored several technical reports, and has conducted formal peer reviews for both research proposals and journal articles. Her relevant background includes 3 years working for the USGS as part of the National Water Quality Assessment Program (NAWQA) and co-leading the development and testing of stream habitat and macroinvertebrate monitoring protocols for the NPS Inventory and Monitoring program's Colorado Plateau Network.

Brett G. Dickson, PhD

Brett is the President, Chief Scientist, and founder of Conservation Science Partners. He is a conservation biologist, landscape ecologist, and ecological modeler with over 25-years of experience in conservation science and project management. He has published over 50 peer-reviewed journal articles or book chapters, each utilizing advanced spatial data analyses for inferences about human impacts to landscapes, animal-habitat relationships, and conservation in the West. Brett also is an Associate Professor at Northern Arizona University, where he co-directs the Lab of Landscape Ecology and Conservation Biology. In these capacities, he has served as a PI or co-PI on over 60 different scientific projects, and frequently works with collaborators and project partners to communicate research outcomes to diverse audiences.

Peer Review Expert Panel

Jesse R. Barber, PhD – Boise State University

Wildlife, Noise

Jesse is an Associate Professor at Boise State University in the Department of Biological Sciences. His Sensory Ecology lab investigates behavioral, evolutionary and conservation-related questions focused on understanding how animals process sensory input and act on the resulting information. One focus of his lab is understanding the effects of anthropogenic and natural noise on ecosystems using techniques spanning laboratory behavioral tests to landscape-scale playback experiments. He has published 30 peer-reviewed papers and currently leads several federally funded research projects.

David Beauchamp, PhD – USGS Western Fisheries Research Center

Water Quality, Fisheries

Dave Beauchamp recently became the Ecology Branch Chief for the USGS Western Fisheries Research Center in Seattle. He has been a Professor in the School of Aquatic and Fisheries Science at the University of Washington and Assistant Unit Leader or acting Unit Leader of the USGS Washington Cooperative Fish and Wildlife Research Unit since 1999. He has conducted research on native nearshore fishes and on pelagic food web dynamics in Lake Tahoe off and on since 1990. In general, his primary areas of research include tactical food web ecology, bioenergetics modeling, and development and application of visual foraging models, all designed to address factors that limit survival and growth of salmonids in freshwater and marine environments. His research program uses a mechanistic approach that scales up from the behavior, bioenergetics, distribution, and sensory capabilities of individuals to the structure and function of aquatic communities. Much of his research focuses on pelagic food web interactions of anadromous and resident salmonids in large lakes and reservoirs, riverine, estuarine, and marine habitats. Common research projects involve determining factors that limit production and survival of key species in response to non-native species, climate change, conflicting demands for water, and other human perturbations. He and his team has conducted numerous quantitative evaluations for the effects of environmental conditions, temporal food supply (i.e., seasonal carrying capacity), competition, and predation on survival and production of salmonids. This approach has been used to examine the feasibility of introducing, re-introducing or conserving anadromous or resident salmonids, Bull Trout, or Kokanee above dams in modified lakes in the Columbia River Basin (Lewis, Clackamas, Deschutes, and Yakima drainages; Lake Chelan, Lake Roosevelt, Flathead Lake), throughout western North America and in the southern hemisphere.

Derek Booth, PhD, PE, PG – University of California, Santa Barbara

Soil Conservation

Dr. Booth is an Adjunct Professor with the Bren School of Environmental Science and Management at UC Santa Barbara and an Affiliate Professor at the University of Washington, and the co-Senior Editor of the international scientific journal *Quaternary Research*. His research focuses on the impacts of watershed disturbance and urbanization on hillslope erosion, stormwater runoff, and stream-channel geomorphology. He holds a B.A. in Geology from UC Berkeley, and M.S. in Geology from Stanford University, and a Ph.D. in Geological Sciences from the University of Washington.

Robert Burns, PhD – West Virginia University

Scenic Resources, Recreation

Dr. Robert Burns is Professor of Recreation, Parks, and Tourism in the WVU Davis College School of Natural Resources. He has conducted research in lake and river-based settings over the past 20+ years. He is Primary Investigator for a long-term US Forest Service (2000-present) visitor monitoring effort (USFS National Visitor Monitoring) in several US western states and the Pacific Northwest (Oregon/Washington/California, Colorado). This social science research effort has resulted in robust trend data regarding visitor use on public lands, with numerous products (peer-reviewed journal articles, book chapters, technical reports, and funding of multiple graduate students). He is the Principal Investigator of a federally funded Cooperative Agreement 14-DG-11132762-156, Amendment 001; *Developing a Student-Centered Visitor Use Management Research Program: ICMBio Parks and Protected Areas in the Amazon Region of Brazil*; agreement expiration date is Dec 31, 2019). He is Chief Editor of the Journal of Park and Recreation Administration and Founding Co-Editor of the International Journal of Outdoor Recreation and Tourism. He is co-editor of the 2016 textbook entitled "*Outdoor Recreation Planning*," published by Sagamore Publishing, and presents papers at numerous symposia steering committees in the US and internationally. Burns is a 20+ year veteran of the US Army.

Stephanie Guildford, PhD – University of Minnesota, Duluth

Water Quality

Dr. Stephanie Guildford is an Associate Professor Emeritus with the Large Lakes Observatory and Department of Biology, University of Minnesota, Duluth. Her research focuses on factors that control algal growth and species composition in large lakes. She has published on large lakes in North America and Africa. She is currently serving as co-editor for the Journal of Great Lakes Research and as Secretary for the Lake Winnipeg Foundation and is involved in research on New Zealand's iconic Lake Taupo. She holds a B.Sc. from Dalhousie University and an M.Sc. and PhD. From the University of Manitoba.

Gary T. Hunt, QEP – TRC Environmental Corps

Air Quality, Noise

Mr. Hunt, a Principal Scientist and Vice President with TRC Environmental Corporation, is an air quality expert with over 39 years of experience. He specializes in the characterization, quantification and control of toxic air pollutant emissions, as well as, their distribution and occurrences in the environment. He is a member and Fellow of the Air and Waste Management Association (AWMA), a member of the American Chemical Society (ACS), Division of Environmental Chemistry and a member of the American Society of Mechanical Engineers (ASME). Mr. Hunt is a Qualified Environmental Professional (QEP) who holds a B.S. in Chemistry from Villanova University and an M.S. in Environmental Sciences from Rutgers University. He was also a participant in the 2012 TRPA Peer Review process. Mr. Hunt has authored more than 100 journal manuscripts and symposia presentations on air quality and other environmental topics.

Jeffrey Marion, PhD – Virginia Tech University

Scenic Resources, Recreation

Dr. Jeff Marion is a scientist specializing in Recreation Ecology, which investigates the environmental impacts of visitor use in protected natural areas. His research has focused on visitor impacts to campsites, recreation sites, formal and informal trails, and climbing impacts to cliffs. He was a founding member of the Leave No Trace

Board of Directors and he chaired the committee that guided development of the Leave No Trace principles, practices, and educational courses for the first decade. He has recently authored the LNT Center's book "Leave No Trace in the Outdoors."

Robert J. Naiman, PhD – University of Washington

Vegetation Preservation, Fisheries

Dr. Naiman is Emeritus Professor at the School of Aquatic and Fishery Sciences at the University of Washington. Career highlights include being a research scientist and director of the Matamek Research Program of the Woods Hole Oceanographic Institution, director of the Center for Streamside Studies at the University of Washington, a visiting scientist on numerous occasions with the Centre National de la Recherche Scientifique (Toulouse, France), and a professor at the Centre of Excellence in Natural Resource Management at the University of Western Australia for several years. His research addresses the structure and dynamics of riverine ecosystems – including riparian vegetation, and the role of large animals in influencing ecosystem dynamics. The research activities laid the foundation for ~11 books on aquatic ecology and watershed management and produced 230+ journal articles. He is an ISI Highly Cited Researcher. In 2008 he was awarded the title of Doctor Honoris Causa by the Université Paul Sabatier, in 2012 he received the Eminent Scientist award from the Ecological Society of America, and in 2013 he presented the E. Baldi memorial lecture to the International Society of Limnology. Until recently he chaired the Independent Scientific Advisory Board for the restoration and management of the Columbia River (USA).

Barry R. Noon, PhD – Colorado State University

Wildlife

Barry R. Noon is a professor in the Department of Fish, Wildlife, and Conservation Biology at Colorado State University. He graduated from Princeton University in 1971 with a B.S. degree in biology and from the State University of New York-Albany in 1977 with a PhD in ecology. In collaboration with many outstanding students and post-docs, he has conducted research on the effects of land management practices on wildlife populations for the past 40 years. His focus has primarily been on the conservation of imperiled species in forest ecosystems. During this period, he has published over 120 scientific papers and co-authored 4 book-length reports to the federal government on the sustainable management of public lands. For 11 years, he directed a Forest Service Research Lab in the Pacific Northwest (USA) and in 1995 served as Chief Scientist of the National Biological Service, Department of the Interior. During the last 15 years, he has served on federal advisory committees providing recommendations to the Secretary of Agriculture on the management of Forest Service lands to better sustain biological diversity and to the Secretary of the Interior on changes to the Endangered Species Act to encourage conservation on private lands. He has also served as chair of the global policy committee for the Society for Conservation Biology and provided testimony to the U.S. Congress on numerous occasions on issues regarding the conservation of wildlife in the U.S. and internationally. Dr. Noon has received several academic awards including the Edward T. LaRoe award from the Society for Conservation Biology (1997), an Aldo Leopold Leadership Fellowship (2004), Colorado State University Distinguished Ecologist (2008-09), and two Senior Fulbright Fellowships to India from the U.S. State Department (2003-04 and 2010-11). In collaboration with his students, his current research focuses on tiger conservation in India, the effects of energy development on imperiled species in the United States, climate change effects on wetland birds, and promoting biodiversity conservation on U.S. Department of Defense lands.

Kevin Rose, PhD – Rensselaer Polytechnic Institute

Water Quality

Dr. Kevin C. Rose is an Assistant Professor in the Department of Biological Sciences at Rensselaer Polytechnic Institute where he holds the Kolleck Chair for Freshwater Ecology. Dr. Rose's research expertise is in aquatic biogeochemistry and limnology, with an emphasis on carbon cycling and water clarity. He has authored more than 30 peer reviewed publications on topics such as water clarity in Lake Tahoe and in journals such as Science. He is also an associate editor for the journal Water Resources Research and is currently editing a special issue for the Global Lake Ecological Observatory Network (GLEON) for the journal Inland Waters.

Scott N. Spak, PhD – University of Iowa

Air Quality

Dr. Spak is an Assistant Professor of Urban & Regional Planning at the University of Iowa. His research uses coupled environmental and human systems models to inform and study local and regional environmental policy. He holds a BA in Engineering Sciences from Dartmouth College and PhD in Atmospheric & Oceanic Sciences from the University of Wisconsin-Madison.

Carol Wessman, PhD – University of Colorado, Boulder

Soil Conservation, Vegetation Preservation

Dr. Carol Wessman is Director of the Environmental Studies Program (Jan 2016), a professor in the Department of Ecology and Evolutionary Biology, and Associate Director of the Ecosystem Science Division in CIRES at the University of Colorado, Boulder. She received a Bachelors Degree (botany) from Colorado State University and Master of Science (remote sensing) and PhD (forest ecology) from the University of Wisconsin-Madison. Her research program emphasizes the dynamics between landscape structure and ecosystem functioning, with focus on understanding ecological response to multiple disturbances. Her approach involves field studies and GIS and remote sensing methodologies investigating temporal and spatial heterogeneity in ecosystem properties. Dr. Wessman's work incorporates theory in ecosystem and landscape ecology, with an emphasis on resilience and adaptation to climate and environmental change. Current research projects include multiple disturbance interactions in subalpine forest; ecosystem resilience; and social-ecological systems in urban environments.

Review by Dr. Jesse Barber

Boise State University

Introduction Chapter: Page 1-5: The lack of inference from regression analysis is perhaps couched more accurately as a limitation in experimental design, not necessarily a limitation in the modeling technique chosen. For example, imagine an experiment where vehicles were not allowed to drive on some sections of road near sensitive wetland areas, during songbird dawn chorus, while other areas were open to vehicles at all times. A regression analysis of the variation in background sound levels created by these conditions and bird reproductive success might reveal a strong relationship. This hypothetical result would have much stronger inference than the description currently in the report attributed to this statistical technique. Thus, I propose, this section should be amended to state the correlational approach to data collection in the report, not necessarily the stats.

Methods Chapter: Overall, very nicely done. The clarity of the status/trends/confidence designation is commendable.

Page 2-10, Table 2-3: Surely there is an error here. Frequentist statistical convention is $p<0.05$ is 'statistically significant'. The R-squared cannot be interpreted if the test is not significant. Do you mean to use these categories: 0.01 and 0.01-0.04 and 0.04-0.05??

Wildlife Chapter: In general, wildlife is well addressed in the 2015 Threshold Evaluation. However, the taxonomic focus, primarily on raptors is perhaps a skewed use of resources. Particularly in light of the fact that the laudable past monitoring efforts have revealed these species are in good shape. It is my opinion it is time for the team to turn their attention to other wildlife groups.

Songbirds and bats are the targets I suggest. Both groups are relatively easy to monitor and hold the potential to reveal ecological patterns of habitat quality broadly (Chapman and Reich 2007; Jones et al. 2009). Both groups strongly depend on insects, are critical biological insect control agents (Kalka et al. 2008; Bohm et al. 2011), and provide an indirect means of assessing insect populations.

As the famous evolutionary biologist, Edward O. Wilson said, 'It is the little things that run the world' (Wilson 1987).

The recent discovery of White-Nosed Syndrome in Washington State (https://www.whitenosesyndrome.org/) should place bats as a top priority for the Threshold Evaluation team. Understanding bat populations now, and the influence of disturbance on their populations, is critical in the short time left before White Nose Syndrome reaches the Tahoe Basin. Bats known to be impacted by this devastating disease are common in the Tahoe Basin (e.g., big brown, *Eptesicus fuscus* and little brown, *Myotis lucifigus*).

I have one comment regarding the management of disturbance free zones: if noise (sound levels) and light at night from human-made sources are not incorporated into what is considered a disturbance free zone the designation 'disturbance free' is inaccurate (Francis and Barber 2013; Gaston et al. 2013).

Major Deficiency: The team needs to address the impacts of roads on wildlife in the Tahoe Basin. The impacts of roads on wildlife deaths from collisions, habitat degradation from noise pollution and, importantly, connectivity are critical gaps in this analysis (van der Ree et al. 2015).

Noise Chapter

Introduction: Page 10-1: Noise influences far more than just wildlife behavior and has been documented to have negative impacts on community structure, distributions, habitat quality and reproductive success (Francis and Barber 2013; Ware et al. 2015).

Page 10-3: Instantaneous thresholds for single events can be useful to enforce and monitor compliance of mitigation measures but integrated sound values over long time periods are the most critical for wildlife and, arguably, for people.

CNEL is not necessarily the best metric to evaluate noise for wildlife, as the penalties are human-centric. However, the penalties of 4.77 dB between 7-10 PM and 10 dB between 10 PM and 7 AM do protect wildlife during biologically critical crepuscular (dawn and dusk) periods.

Further, A-weighting of sound levels is an acceptable metric for humans and birds but is not adequate for many wildlife species (Francis and Barber 2013). Flat, un-weighted decibel readings should be used instead.

The Natural Sounds and Night Skies Division of the National Park Service has a defined protocol to monitor sound levels in natural areas, has the longest history of any land management group of monitoring sound levels and could be a valuable resource for the Threshold Evaluation team when addressing this ecological pollutant.

Thresholds: Critically, there is little to no evidence that the threshold of 45 CNEL is fully protective of the most critical wildlife habitat. Many animals have hearing abilities that are many orders of magnitude below our own, and many natural habitats experience low-term integrated sound levels far below 45 dBA, often near 0 dBA. I suggest that Critical Wildlife Habitat Areas and Wilderness and Roadless Land Uses designated land types be maintained at the average sound levels of areas that currently experience the least amount of anthropogenic noise exposure. Ideally, these critical areas would be maintained at their natural ambient levels, however given the prevalence of airplane traffic it is unlikely that any habitat currently experiences its ancestral sound levels. It seems clear that any increase in background sound level beyond ambient from natural sounds is likely to change ecological functioning (Shannon et al. 2015). For the most important habitat, I suggest using the precautionary principle.

Page 10-18. Saying there is insufficient data to determine the trend seems pretty erroneous to me here. Levels are over threshold. Traffic has increased since 2013. The habitat is still over the threshold. It should be managed as such.

Methods: How was the sound level data collected (equipment, protocols, etc.)? If done according to best practices, all metrics (LEQ as opposed to weighted CNEL; dBF as opposed to dBA; time above metrics, etc.) should be able to be calculated. These details need to be included.

References

Bohm SM, Wells K, Kalko EKV (2011) Top-down control of herbivory in birds and bats in the canopy of temperate broad-leaved oaks (Quercus robur). *PLoS One* e17857.

Chapman KA, Reich (2007) Land use and habitat gradients determine bird community diversity and abundance in suburban, rural and reserve landscapes of Minnesota, USA. *Biological Conservation* 135(4):527-541.

Francis CD, Barber JR (2013) A framework for understanding noise impacts on wildlife: an urgent conservation priority. *Frontiers in Ecology and the Environment* 11(6):305–313.

Gaston KJ, Bennie J, Davies TW, Hopkins J (2013) The ecological impacts of nighttime light pollution: a mechanistic appraisal. *Biol Rev* 88(4):912–927.

Jones G, Jacobs DS, Kunz TH, Willig MR, Racey PA (2009) Carpe notem: the importance of bats as bioindicators. *Endangered Species Research* 8:93-115.

Kalka MB, Smith AR, Kalko EKV (2008) Bats limit arthropods and herbivory in a tropical forest. *Science* 320(5872):71.

Shannon G, McKenna MF, Angeloni LM, Crooks KR, Fristrup KM, Brown E, Warner KA, Nelson MD, White C, Briggs J, McFarland S, Wittemyer G (2015) A synthesis of two decades of research documenting the effects of noise on wildlife. *Biological Reviews* DOI:10.1111/brv.12207.

van der Ree R, Smith DJ, Grilo C Eds. (2015) *Handbook of Road Ecology*. Wiley and Sons. 552 pages.

Ware HE, McClure CJW, Carlisle JD, Barber JR (2015) A phantom road experiment reveals traffic noise is an invisible source of habitat degradation. *Proceedings of the National Academy of Sciences* 112(39):12105–12109.

Wilson EO (1987) The little things that run the world (The importance and conservation of invertebrates). *Conservation Biology* 1(4)344-346.

Review by Dr. David Beauchamp

USGS Western Fisheries Research Center

Introduction

In general, this section is clear and informative. Just a few specific comments follow.

History

P 1-1 Instead of simply referring to the Comstock era, please provide a range of dates to bracket the specific periods when timber and grazing impacts degraded habitat within the basin.

Similarly, provide the year(s) when forest management policy of fire prevention and vegetation preservation policies were endorsed.

P 1-24th paragraph: "Since the 1960s,…" Please provide the actual population figures during the 1970s.

P 1-9 Indicator: replace "constituent" with a more meaningful term. Most indicators should be in the form of a "variable." "Parameters" are highly mis-used terms in limnology and other fields, and are often variables that are mis-labeled as parameters.

Methods

I appreciated the thorough definitions and derivation of key calculations in this section. This section will pre-empt needs for further clarification throughout the documents that would have otherwise emerged.

P 2-1 "Environmental Threshold Carrying Capacity (Threshold Standard)" Delete "Carrying Capacity" and simply refer to "Threshold Standard" instead. The term "carrying capacity" has very specific meanings depending on context, and could lead to unintended interpretation.

P 2-10, paragraph below Table 2-3 beginning with "Where the data collected…" Something is missing or poorly phrased in the second line: "…and the test for statistical [… revealed both indicate]…

Chapter 4 Water Quality

In general, there was considerable redundancy throughout this chapter, because sections on nutrients and turbidity were divided into separate treatments of: 1) concentrations, and 2) total loading. Since total loading requires knowledge of daily concentrations x flow, combining these metrics into the same section(s) would provide more context for interpreting when high concentrations should be a real concern and could help identify and prioritize future restoration efforts. For instance, we should worry less about high concentrations of nutrients or turbidity when associate with relatively low flows or very short durations of high-concentration events than when associated with higher flows or more prolonged events. The highest total loading of both nutrients and turbidity typically come from watersheds with moderate concentrations, but higher total flows. So placing the concentration data in the context of total flows and thus total loading seems like a more meaningful way to present these data.

Unless this format is obligatory for some reason, I strongly recommend that the sections on concentrations and total loading be combined, then simply highlight the different implications and interpretations produced by these metrics. Some tables could simply add 1-2 more columns to combine the concentration and total loading

data into a more synthetic and accessible form. Combining these sections would also reduce the length of this chapter significantly (perhaps by 30-40%).

P4-2. Equally important is the exceedingly small areal extent of the watershed relative to the size of the lake.

Some acknowledgement of how the loss of Daphnia after the Mysid introduction affects lake transparency would be important. The much higher grazing capacity of Daphnia compared to copepods, etc. and vertical distribution patterns of Daphnia (primarily epi- and meta-limnetic) often combine to significantly increase transparency for lakes of similar trophic status. The loss of this relationship should also be highlighted in the Fisheries Chapter, because it was both a more efficient link from primary production to fish production, but also as a process that can improve transparency.

Table 4-1. Are macrophytes covered in the Littoral Invasives section or elsewhere?

Macrophytes are an important concern for littoral dynamics and species composition and interactions. It would be helpful to explicitly list at least some of key groups of invasives intended here.

Table 4-2. Again, monitoring recent and existing aquatic invasives in both pelagic and littoral habitats should be assessed on a routine basis (e.g., during key month or months every 1-3 years) to facilitate early or rapid shifts in status that could generate detrimental impacts on water quality, food web stability, etc.

Too many table entries list insufficient data to even determine a trend!

Lake Tahoe Pelagic Waters

P 4-18: To clarify the definition of pelagic habitats in the 2nd sentence, please revise: "… all waters of the lake [with bottom depths] deeper than 30 feet or…."

P 4-20-22. Trends in Secchi depth transparency are encouraging, even after accounting for potential bias from recent drought conditions.

P4-22. The 24,644 miles of street sweeping is presumably achieved by multiple sweeping over the same roads. Please clarify, e.g., sweeping on a regular cycle? Only where and when as needed?

P4-25. Monitoring Approach. Presumably For practical reasons, the photic zone has been defined as some fixed range of depths (based on previous monitoring) from which depth-specific samples are drawn for estimating depth-integrated ppr, or are potentially different depths sampled every trip, depending on optical characteristics of that day? If a fixed depth is used, then please clarify by stating that:

"…water samples are collected from [the surface to xxx m at] 13 different depths spanning the photic zone [based on ???]."

Alternatively, if sampling depths are adapted to the photic zone estimated for that day, please explain the process

Tributary Water Quality

Suspended sediment concentration threshold (60 mg/L).

P4-55. Midway through the monitoring Approach box, Please clarify that each stream averaged 28 sampling events per year: "For the last 10 years, the average number of samples collected [annually was N=28 for each stream]…"

Note-The phrasing for the average number of samples taken annually per stream varied throughout the document. This particular example of phrasing was quite ambiguous whereas most others were better. It'd be useful to use clear consistent phrasing throughout.

P4-59. The recommended alternative analyses that would use time- and flow-weighted daily averaging for suspended sediment/turbidity and nutrients should be adopted wherever sufficient data are available.

The potential day-night sampling bias ("time-of-sampling bias") is a legitimate concern and measures to address this should be prioritized.

P4-63. The different phosphorus standards used by Nevada (0.05 mg/L) and California (0.015 mg/L) should be consistently footnoted whenever status or trends are reported in Tables or figures.

Relevance—

-The regression of Total P versus turbidity would provide valuable quantitative support for the the degree to which TP affects lake transparency

-I'd prefer that the value of algae to the aquatic food web be stated a bit more broadly (i.e., algae directly feeds primary consumers which are then consumed by invertebrates and vertebrates at higher trophic levels)

Monitoring and Analytic Approaches: Again, combining the concentration and total loading sections would be very helpful here. When dealing with the concentration data in isolation, I started to be distracted by the concern that the mass-balance calculations for total loading would be compromised by how "average annual concentrations…" were computed, whereas if these sections were combined, the text would be significantly shoter, clearer, and much more informative.

P4-65 Another example of: "For the last 10 years, the average number of samples collected [annually was N=28 for each stream]…"

P4-68. Analytic approach. Consider adopting the time- and flow-weighted averaging for TP. Again, this argues for combining the concentration and total loading sections together!

Monitoring approach-if sufficient data are already available for some tribs, regression equations for TP versus turbidity would be valuable.

Nitrogen Concentration

Footnote the different TN concentration standards for Ward and General (0.15 mg/L) versus Blackwood, Trout, and Upper Truckee (0.19 mg/L) whenever referring to proportion of samples exceeding the standards, or for status and trend summaries.

P4-71. Again-please broaden the ecological contribution of primary producers to higher trophic levels in aquatic food webs (not just to primary consumers).

P4-72. Great to have cross-calibrated 5.5 years of nitrate + nitrate data using old and new methods.

Tributaries Suspended Sediment Load

P4-78 Adopted Standards-Please clarify standard 2) Littoral and pelagic Lake Tahoe: 3 NTU is quite a lot of turbidity for the pelagic zone. Please add more description as to whether this 3 NTU value pertains to individual measurements or some broader average across dates and/or locations.

P4-79. Analytical approach. Seems like many of the same analytical approaches were used for suspended sediments, phosphorus (TP) and nitrogen (TN). Can't these methods be combined in a much more efficient way for shared methodologies, then simply highlight differences specific to a particular constituent?

P4-83. Effectiveness of Programs and Actions. A recurring statement that the relative importance of specific factors cannot be dissected out because so many factors are co-mingled is not very satisfying. Improved analytical methods that can identify and quantify the relative contributions of various factors should be prioritized over the intervening period before the next Threshold Evaluation. Such an effort would produce

tremendous benefits in terms of diagnosing the most problematic factors and would inform how to prioritize restoration efforts more effectively.

Monitoring Approach-It's incredibly important to assess the effectiveness of restoration projects, yet very little effort has been allocated to such assessments historically. I applaud the authors' recognition this important issue. More thoughtful pre-post evaluation of restoration projects will be required to guide us to truly effective restoration practices.

Total Nitrogen Load.

The time series graph of Total N Load and Total Flow shows a phenomenal fit (nearly 1:1). Is this real? Earlier statements about how microbial action decoupled nitrogen availability from loading seems to be contradicted by this figure.

P4-104. Modification of the threshold standard or indicator-

-Need to develop tributary load reduction targets in order to establish some benchmark for evaluating status, trends, and progress toward attainment. These targets can be modified adaptively as new information emerges, but in their absence the monitoring, evaluation, and restoration efforts will be unnecessarily open-ended and lack effective guidance.

Surface runoff

P4-109. Effectiveness of programs and actions. With regard to the diminishing returns from increasing storm water retention beyond the 20-year 1-hr storm. While valuable to evaluate potential cost-benefit ratios, stating just the average response curve seems overly simplistic here. I doubt that all locations behave so similarly. An incremental change in retention capacity at some locations could produce much greater (or lesser) benefits than others. Does the citation (2nd Nature & NHC 2011) show the variability in responses among locations or conditions? If so, are the responses similar enough to support the claim that a 20-yr 1-hr retention capacity is a good all-around standard? If not, then a bit more detail would be really valuable here.

Implementation and Effectiveness-

P4-114. If only 186 BMP certificates have been issued out of 2441 parcel owners that were notified that maintenance was due, then doesn't that imply only 7-8% compliance? Sounds like an enforcement issue here.

P4-125. Recommendations-Monitoring approach.

Some good points listed here. They reinforce the suggestion that monitoring and analysis for concentrations and Total annual loading should be combined from the sampling design through final analysis phases.

Chapter 7 Fisheries

P 7-1, 2nd paragraph: A Table of native and non-native fishes, relative abundance (extirpated, rare, common), population trends (increasing, declining, stable) and approximate dates of discovery or extirpation would be useful.

Including 3 key invertebrates in this table (Daphnia, Crayfish and Mysids) would be helpful as well.

P 7-2, Policy points:

2. Some stream blockages may be beneficial for inhibiting the spread of non-native species, so perhaps include some caveats to acknowledge the potential usefulness of some blockages as a conservation tool.

5. Encouraging [functional] habitat improvement projects in streams and lakes.

7. Clarify by rephrasing (The intent of this point was initially confusing). Proposed edit: “Transferring existing water diversions from streams to lake withdrawals, whenever feasible.”

P 7-3, Table 7-1 Lake Habitat:

The essential habitat paradigm developed in stream ecology is only partially effective in lake environments, especially in large lakes. Biotic interactions (food supply, predation, competition) become much more important processes regulating lake communities than simply physical habitat alone. Threshold standards should be expanded accordingly to ensure that appropriate monitoring metrics are tracked to evaluate status, trends, and key processes that underlie changes in health of fisheries resources.

Pelagic habitats cannot be ignored. Pelagic habitats support native Lahontan redsides, Tui chub, (and probably larval sculpins), and historically supported Lahontan Cutthroat trout.

These habitats have been changed by non-native species (mysids, lake trout, etc.), loss of Daphnia, and warming thermal regime

The fisheries section should develop additional threshold standards to include:

-Pelagic habitat standards (density of adult crustacean zooplankton, timing of thermal stratification, and the depth and temperature of epilimnion during summer)

-Non-native spp standards (relative abundance, spatial distribution, size structure)

Current lake habitat indicator (acres of nearshore habitat defined by substrate size): Does this account for seasonal hydrodynamics that would naturally sort and maintain different substrate size distributions along different shorelines? I worry about potential efforts to add or change substrate composition. These will not be sustainable unless they operate in concert with existing physical processes of wind, wave energy, water circulation, and sediment accretion or erosion zones.

Stream Habitat: Does this indicator factor in the need for connectivity among essential habitats that might support different life stages during different seasons? Some stream miles are more influential than others. Otherwise valuable habitat might not be useful if too fragmented or isolated from other reaches.

Instream Flow: The 2nd standard is confusing. Please clarify-is the intention to transfer water withdrawals from stream diversions to direct withdrawal from the lake? Phrasing for this in Table 7-2 is more informative: "Divert stream intakes to lake sources"

Lahontan Cutthroat Trout: Adfluvial Lahontan cutthroat trout will require good pelagic conditions to support adequate growth and survival. This is more of a comment to highlight the need to increase the “Lake Habitat Standards” above.

Lake Habitat

P 7-5: 1st paragraph. “Feed and cover habitats” It will be important to differentiate habitat needs of larval forms of minnows and suckers, which have more limited ability to move, feed, and deal with wave energy, from post-larval "fry."

I don't believe that enough is known about essential larval fish rearing habitat, but I suspect that the lower wave-energy sites with smaller substrate, might be very important for larval stages, but less important for fry and older juveniles.

3rd paragraph: The decline in native fishes is likely more of a response to interactions with non-native species than change in littoral habitat alone. Reproduction, recruitment, growth and survival of fish species can be strongly influenced by seasonal food supply, thermal and other environmental conditions, interactions with native and non-native species; all of which can fundamentally change whether the existing physical habitat can

adequately support these populations as they might have historically. Thus the need for additional threshold standards for lake habitat that relate to non-native spp, ecological and environmental conditions that affect survival and growth of native fishes.

Direct monitoring of trends in native and non-native fishes (and key invertebrates like crayfish and mysids) should be included as indicators. For native minnows, quantitative, depth-stratified minnow trapping would be an effective low-cost approach for monitoring status and trends of these populations, and would benefit from existing time series data.

P 7-8 & 7-9 Indicator State:

Efforts to "restore 'prime' habitat" would be wasteful in the absence of evidence that habitat is the primary limiting factor responsible for suppressing fisheries resources.

Instead, improved information on which habitats support successful spawning-larval-fry transitions, monitoring non-native species and their effects on native fisheries resources are more pressing actions.

"…additional factors influence the quality of littoral fish habitat."

YES-monitoring of native and non-native species (relative abundance, distribution, size structure) should be integrated into this threshold standards framework.

Trend: I agree that the methodological differences employed through time to assess littoral habitat represent a refinement in habitat mapping rather than a real change in habitat quality/quantity.

Again, directly measuring trends in fish distribution, relative abundance, and size structure should receive more attention rather than a sole focus on physical habitat.

Authors report High confidence in a meaningful trend (reduction in native minnows), yet this metric is not currently recognized as threshold standard for fisheries. This should be changed.

P 7-9 Implementation and Effectiveness: Does habitat mitigation occur at the site of disturbance, or can it be targeted for high-value restoration/mitigation actions elsewhere in the basin? Please clarify.

Effectiveness of Programs and Actions: While prevention of new invasive species introductions is unquestionably important, the more critical issue is whether some existing non-native species are expanding abundance, distribution, or impact on native species.

Recommendations:

Additional the additional of more direct threshold standards for status and trends of littoral fishes and pelagic habitat thresholds should be adopted as mentioned in sections above. These additions would then require additional monitoring and analysis beyond the current focus on littoral habitat.

Modification of Threshold Standard or Indicators:

I agree that the 2016 map be adopted. Remote sensing technology and analytical capability will naturally continue to rapidly improve. So reduce concern about future quality of this metric-no need for aggressive efforts to improve.

Adopting a formal monitoring program for assessing status and trends of native littoral species and non-natives is essential. As repeated from sections above Additional Threshold Standards and Indicators should include:

-Pelagic habitat standards (density of adult crustacean zooplankton, timing of thermal stratification, and the depth and temperature of epilimnion during summer)

-Non-native spp standards (relative abundance, spatial distribution, size structure).

Stream Habitat

7-13 Devote a whole page to this map and label the tributaries. It's virtually useless at this small size and poor resolution.

7-14 Monitoring and Analysis:

MMI-As a key metric for evaluating stream habitats, this index should be described in moderate detail in the Methods Chapter. Specifically, what measures of “ecological structure and function” are included, and at what temporal-spatial scales are they measured?

What I'm NOT seeing here is any direct correlation to fish production, growth, survival, size structure, etc.

Reliance on indirect measures of "stream health" doesn't necessarily translate to health of fish populations.

Because continuous temperature data, fish sampling data, etc. were not explicitly included in threshold standards for stream habitat condition, I'm concerned that this assessment will be misleading with regard to status and trend of fisheries

P 7-15. Top paragraph. Assigning a single rating to an entire stream is a real limitation, especially for the major drainages.

P 7-16. 22 degrees C is nearly an acute lethal temperature for most salmonids. A more reasonable threshold for chronic stressful exposure would be 18 degrees C. Cite some peer-reviewed literature here from Dan Isaak (USFS-Boise Experiment Station) and colleagues:

Wenger et al. 2011. Flow regime, temperature, and biotic interactions drive differential declines of trout species under climate change. Proceedings of the National Academy of Science www.pnas.org/cgi/doi/10.1073/pnas.1103097108

Jager HI, Van Winkle W, Holcomb BD (1999) Would hydrologic climate changes in Sierra Nevada streams influence trout persistence? Trans Am Fish Soc 128:222–240.

McCullough DA, et al. (2009) Research in thermal biology: burning questions for coldwater stream fishes. Rev Fish Sci 17:90–115.

Sustained thermal tolerance is also strongly mediated by food supply: greater food supply enables greater tolerance to warmer conditions or longer periods of exposure.

A more explicit description of temperature monitoring (locations and whether continuous logging versus spot measures) is needed here.

Good physical habitat features won't support salmonids if temperatures are too high. Invertebrates do not necessarily share the same thermal tolerance or performance responses as fish, so BMI is also not a great correlate for stream habitat quality for fishes.

Implementation and Effectiveness. Effectiveness of Programs and Actions:

P 7-18 Annual monitoring should be adapted to directly evaluate responses to restoration projects. What's effective and what isn't? Unless restoration projects can demonstrate measurable and sustainable benefits, more of the same cannot be justified.

A bit more quantifiable statement here would help. The implication above that cumulative effects will eventually provide measurable benefits needs a comprehensive vision for how connectivity of quality habitat structure will improve the function and carrying capacity of stream habitat to support fisheries resources. What is the vision for achieving a meaningful "critical mass" of habitat restoration?

P 7-19. Recommendations:

Without a more direct assessment of fish status, trends, or mechanistic examination of relevant limiting factors for native fishes, the monitoring and analytical approach are inadequate.

The current approach essentially tracks performance of benthic invertebrates and physical habitat features while ignoring or at least inadequately including more directly relevant metrics like diel-seasonal thermal regime and direct measures of size, condition, and relative abundance of fish. Not even presence/absence of fishes was mentioned here.

My apologies if I'm missing something here, but if so, then a bit more detailed description of how fish performance is measured or how fish performance connects directly to the metrics examined here would be very helpful.

Instream Flow

P 7-23 top of page: Related to Tracy & Rost's (2003) evaluation…establish minimum instream flow conditions."

Would be very helpful to link these findings and conclusions to criteria for selecting stream restoration projects, i.e., habitat restoration should only be emphasized in reaches with high probability that adequate flows and temperatures will be available.

Recommendations

Regarding Modifications of the threshold standard or indicator:

Agreed, however, a stronger link between Tracy & Rost's conclusions and how these inform habitat restoration priorities would be valuable here.

Lahontan Cutthroat Trout

Background:

Human & Environmental Drivers: Mysid shrimp should also be listed explicitly here.

Implementation and Effectiveness:

"…continuing challenges include adverse interactions with non-native species…"

Mysis diluviana are a fundamental contributor to change in food web structure and competition for zooplankton, as well as a key energy source that supports lake trout production.

"…additional research is needed to improve understanding of reintroduced LCT population dynamics and their interactions with nonnative species (Al-Chokhachy et al. 2009)."

Understanding how LCT will seasonally utilize habitats and depths, grow, and withstand predation by lake trout are logical next steps in Lake Tahoe.

References:

Missing authors:

Al-Chokhachy, R., M. Peacock, L.G Heki, and G. Theide. 2009. "Evaluating the Reintroduction Potential of Lahontan Cutthroat Trout in Fallen Leaf Lake, California." North American Journal of Fisheries Management 29: 1296–1313.

Review by Dr. Derek Booth

University of California, Santa Barbara

REVIEW OF INTRODUCTION AND METHODS

INTRODUCTION

Overview

This is a clearly written overview of the Tahoe-area history and the overall approach being taken in the rest of the document. Its level of detail seems about right, and the introduction to the indicator icons is clear (although the distinction between "Threshold Category " and "Threshold Indicator Reporting Category" seems unintuitive and possibly unnecessary). Including an evaluation of confidence, and an option for compiling recommended changes to the indicators, is a welcome acknowledgment of the need to embrace uncertainty and ongoing improvements in approach.

Specific comments

Figure 1-1. This map has a date of 2012. It should be clear whether it is "active" or just a handy, available graphic. Note that elements are not legible at page scale and should probably be enlarged or omitted for clarity. The "Stream Restoration Plan Area" for South Lake Tahoe is particularly difficult to interpret; "Mixed-Use" in the legend should probably be "Mixed Use".

Page 1-3. "Threshold standards set environment quality targets to protect the unique natural values of the Tahoe Region while still providing for appropriate and orderly development." This should probably be revised to read "Threshold standards set environment quality targets INTENDED to protect…" There's nothing in this document to suggest that anyone has really evaluated whether these targets, set back in the 1980's, can actually achieve their stated goals. Based on the response-to-comments in the 2011 Threshold Evaluation, that is *not* the task of these documents ("the primary purpose of the threshold evaluation report [is] evaluating status attainment…" [p. E-31 of that review]). This response to the comment raises an obvious question: if not part of the Threshold Evaluation Report, where is this being done?

Because of this apparent shortcoming, the following statement on p. 1-4 does not appear to be supported: "The periodic threshold evaluations reported on progress in achieving threshold standards and put forward Regional Plan course corrections in response to best available science and monitoring." Perhaps greater emphasis should be placed on the stated purpose of the Threshold Evaluation, "Provide recommendations on additional actions to facilitate threshold standard attainment o*r otherwise improve the effectiveness of the plan or applicable standards*" (emphasis added). In other words, this should not just be an evaluation of progress towards a pre-established target, but whether the target itself is appropriate. That's what "best available science" is supposed to inform.

Page 1-5. The limitations of a status-and-trends monitoring program to determine causality is much appreciated—it is a commonly forgotten truth.

Page 1-6. The definition of numerical standard is missing the words "intended to" or "assumed to" or "hoped to" somewhere in its text. Presenting these numerical standards as rock-solid, precise thresholds of certain scientific validity provides no service to the public or managers. Does anyone believe that the provided example ("the annual average deep water transparency as measured by Secchi disk shall not be decreased below 29.7 meters") will protect the stated values at a qualitatively different level then, say, 29.6 meters? Or 29.8 meters? Although the selected value is not arbitrary, nor does it provide three-digit precision for assuring certainty of goal attainment once reach. After all, in some cases such thresholds are later found to be simply wrong.

The text for management standards is more circumspect ("intended to").

For a *policy* statement to provide "specific direction" to an agency (or anyone else) is an oxymoron—policies are normally broad, aspirational statements to frame an overarching intent or goal. Reading through the provided text that follows is a perfect example of what is *not* specific.

METHODS

Overview

Having a centralized text that provides common definitions and approaches, and that outlines the data sources for all subsequent chapters, is a useful structure that goes far to streamline the presentation of the chapters that follow. The risk, however, is that it requires a standardized framework that may not be equally suited to every topical evaluation. Some of this difficulty is suggested by the specific thresholds and boundaries that are used to define the status of attainment or the magnitude of a trend. Avoiding arbitrary or inconsistent definitions from one chapter to the next is certainly commendable, but as noted in the peer review of the 2011 Report not every fractionally equivalent numerical change has the same meaning, or the same consequences, across all resource indicators.

Specific comments

Page 2-1. Trying to set threshold standards that will be necessary (and sufficient) "…for the maintenance of a significant scenic, recreational, educational, scientific or natural value of the Region, or to maintain public health and safety within the Region" is a daunting task, and there should be a high degree of scientific skepticism that the Tahoe Region (or anyplace else) can get them right on the first attempt. However, I see little effort in either the responses to the 2011 peer review or the 2015 chapters that I am reviewing that such an introspective evaluation has been accomplished. I see changes at the margin (e.g., substituting a more current soils map for the original data source) but no substantive critique of the topic of this paragraph, namely the threshold standards themselves.

I also don't see any mention of the balance between environmental quality and economic opportunity invoked on page 1-3 of this 2015 draft ("…while still providing for appropriate and orderly development"). This definition of Threshold Standards is all about natural attributes—have we decided to ignore the economy, or is that an inconvenient (yet inescapable) reality best acknowledged as infrequently as possible?

Page 2-8. I doubt that the "estimated attainment date" will have much meaning for most indicators. Unless they're tied to a concrete, funded plan or program, it's almost silly to use what I assume would be a linear extrapolation of prior trends to "calculate" a date. Absent firm legally and/or financially obligated commitments, just don't do it.

Page 2-9. I would trust professional judgment of confidence ("weight of evidence," if you prefer) much more than a 2-out-of-3-is-good approach. For example, a poor protocol will yield useless (i.e., "low confidence") data no matter how and well-located and frequently they have been collected. That bad (or the wrong) data are collected with high precision and frequency, and with a certainty of where they are located, will not improve their value in assessing status.

Similarly, a mechanical stratification of r^2 values for trend confidence is very limited. P-values aren't much better (see Head et al., 2015, The Extent and Consequences of P-Hacking in Science: PLoS Biol 13, e1002106). Plot (and display) the data, and draw whatever qualitative conclusions you believe are warranted from such a display. I suspect, however, that only rare settings will provide you with sufficient information to quantify an assignment

of confidence with the precision implied by this approach (and r^2 values almost certainly won't provide much basis in any case).

In summary, I see no defensible value in either Table 2-3 or Table 2-4. Please discuss this matter in depth with statisticians before pressing forward in this fashion.

REVIEW OF CHAPTER 5, SOIL CONSERVATION

Overview

The Soil Conservation chapter begins with a laudable affirmation of the importance of protecting soil resources throughout the basin. Articulated benefits include sustaining vegetation, maintaining water quality, providing habitat, and providing "a platform for development." Evaluating the state of such a critical watershed and ecosystem component is based on two reporting categories: impervious cover as a function of soil/landscape position, and stream environment zones (SEZ's).

As noted by a prior reviewer of the 2011 Threshold Evaluation, restricting the scope of this chapter to just two categories of indicators (and, by inference, two sources) of soil erosion appeared rather limiting. The response to this comment by the chapter contributors (Appendix E, p. E-31) stated "The chapter focuses on the impervious cover and SEZ indicator reporting categories *to address the primary purpose of the threshold evaluation report in evaluating status attainment*" (emphasis added). Although administratively defensible, this points to a fundamental shortcoming of this review process—and, by association, this 2015 draft Threshold Evaluation. If the overarching task of the report is not to evaluate the status and trends of Lake Tahoe for present and future generations, but instead to check off a bureaucratic box of potentially questionable relevance or value (see below), then the need for a peer review such as this one is quite limited.

Detailed comments

As explained in the 2015 text (p. 5-1), "Impervious cover is a primary indicator of land disturbance. Excessive impervious surface contributes to sediment and nutrient inputs to Lake Tahoe and its tributaries impairing water quality, altering surface hydrology and groundwater recharge regimes. The results are often negative impacts on soil health, fisheries, wildlife habitat and vegetation growth. SEZ's provide a variety of critical services in the basin…" Although related to some extent, these two classes of indicators will be considered separately in this review.

Impervious surfaces

In every iteration of the Threshold Evaluation report readily available on the internet (2001, 2006, 2011, and now 2015), the use of impervious cover and the associated criteria and thresholds are cloned from that of Bailey (1974, Land-Capability Classification of the Lake Tahoe Basin, California-Nevada, A Guide For Planning. USFS, USDA, TRPA, South Lake Tahoe, CA). Thus, any review of the Soil Conservation chapter of the Threshold Evaluation must begin with a review of this 1974 report. Although the 2011 peer reviewers apparently lacked access to this document, it was readily available as of 2016. It appears to have been produced as an agency report; no record of an externally peer-reviewed document (e.g., a scientific journal article) has been found, and a subsequent update of the data sources and acreage tallies (Loftus 2007) strongly suggest that even internal agency review prior to its publication was minimal.

Bailey (1974) is, as its title clearly states, primarily a "land-capability classification of the Lake Tahoe Basin." It followed a well-described procedure for using map-based data on topography, soil type and texture, and geology and geomorphology to define three hazard "potentials" (Bailey 1974, Table 4; reproduced as TRPA 2015, Table 5-2):

- Slope potential (divided into gradient categories of 0-5%, 5-9%, 9-16%, 16-30%, >30%)
- Erosion potential (based on aggregate stability and soil permeability)
- Runoff potential (based on hydrologic soil group)

This conceptual approach is well-supported in the literature of the 1960's and 1970's, and it bears great similarity to other such efforts of similar vintage. For example, the "Universal Soil Loss Equation" (first released in 1965, updated in 1978) predicts the potential erodibility of a soil at particular location on the landscape by the following parameters:

- Typical rainfall intensity
- Hillslope gradient
- Length of contributing hillslope
- Aggregate stability and soil permeability
- Soil texture
- Vegetative cover

In Bailey (1974), the discrimination of soil and landscape types is followed by a less well-defined process that integrates these potentials into 7 "Capability levels" (actually, 9 such levels because Level 1 is further subdivided into 1a, 1b, and 1c on the basis of soil saturation and "fragile flora and fauna"). Bailey (1974, p. 18) states that "No absolute evaluation of hazard was attempted; only relative hazard within the area was considered. The land capabilities map combines the data presented on each of the other maps to provide a single hazard rating of the basin." However, each hazard rating *is* ultimately quantified precisely, by assigning a discrete numerical threshold for "disturbance" as measured by the fractionally recommended maximum impervious area coverage. This quantification is summarized in a table (Bailey 1974, p. 24) that is reproduced as part of Table 5-1 of TRPA (2015).

The sources of the assigned "allowable percentage of impervious cover" values in Bailey (1974) are asserted to comprise recent local erosion-rate studies, field observations, conversations with others, and a host of studies published in the late 1960's and early 1970's on sediment yields and runoff quantities from urban and suburban areas around the country. No specific values of imperviousness drawn from these studies are cited as a basis for the allowable percentages, and having personal familiarity with most of the cited literature I suspect it is because they do not actually provide any.

As noted in the 2015 draft report, the original soils mapping used by Bailey was subsequently updated by Loftus (2007). Curiously, Loftus's update to Bailey's Table 4 (Loftus 2007, Table 4a) is not included in the 2015 draft, which instead reproduces Bailey's original Table 4 (TRPA 2015, Table 5-2). The 2007 update, however, only reconciles some prior inconsistencies and puts forward a more recent, higher resolution map on which to base the boundaries of different soil categories. It does not re-evaluate the basis for the capability levels, or the specific values chosen, at all.

Although beyond the scope of this review to provide a full critique of this 42-year-old technical report, since transformed without substantive modification or critique into management policy for one of the greatest natural resources of the continent, several aspects of that foundational study bear specific comment:

- *Sediment sources and erosion processes in the Lake Tahoe watershed*

Three fundamental assumptions underlie both Bailey (1974) and all available iterations of the Threshold Evaluation:

1. There are only two significant sources of management-responsive sediment to Lake Tahoe: urban development and degraded riparian zones.
2. The primary mechanisms by which urban development results in sediment production and delivery are surface runoff production, which can result in sheetwash and gullying, or channel erosion from increased discharge in downstream receiving waters.
3. Increased sediment production from degraded riparian zones is a consequence of bank erosion and/or channel incision.

These assumptions are plausible, but none appear to have been subject to even casual inspection, never mind rigorous evaluation, in the context of the Threshold Evaluation. Regardless of what studies have subsequently been conducted in the Lake Tahoe basin over the last 40+ years, their unmodified acceptance from Bailey (1974) suggest that any such evaluations that *have* occurred have not informed the acceptance of these assumptions in their original form.

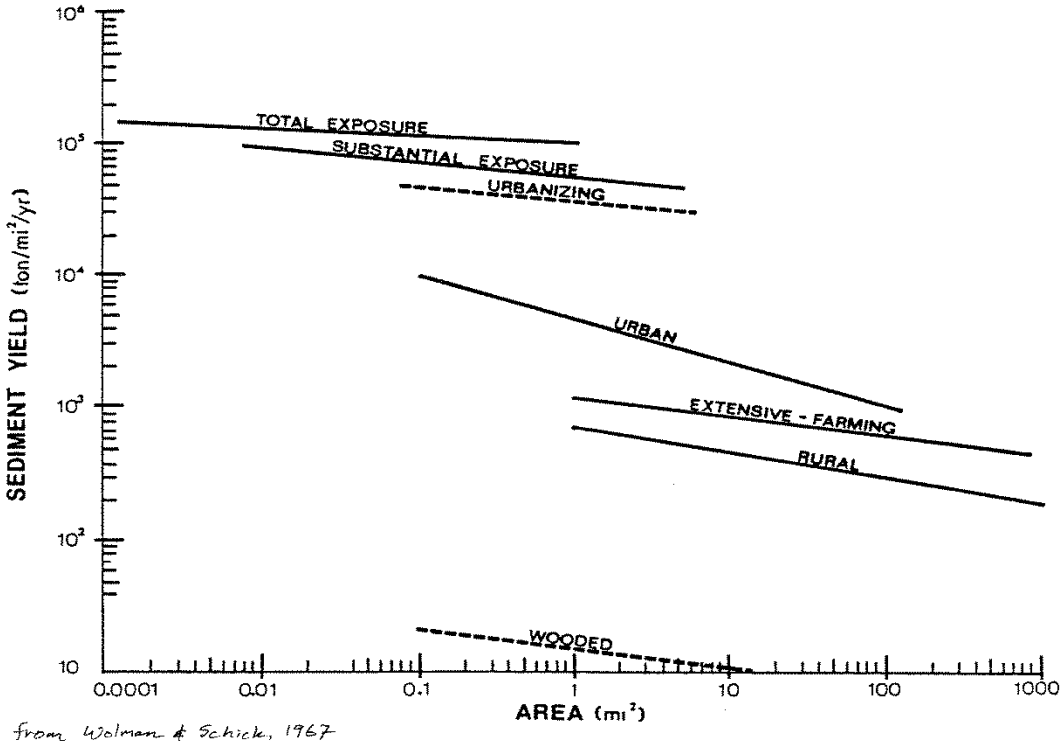
Even a cursory review of recent published studies suggests that much progress has been made on these topics, even if not incorporated into present management approaches. For example, Rios et al. (2014) monitored small urban catchments in the Lake Tahoe Basin and concluded that "small, urbanized watersheds and intervening zones are disproportionately important contributors of nonpoint source pollution, including nutrients and suspended particles." This supports the first assumption listed above; however, they also recognized that impervious area explained only "35.7%" of the variance in suspended sediment yields in their data set, suggesting that this parameter is likely not the major (never mind the only) determinant of fine sediment delivery into the lake, even from the land uses most likely to show such a dependency.

Simon (2008) made direct measurements of streambank erosion, extrapolated those results to the remaining unmonitored drainages across the basin, and compared the resulting totals of measured+inferred fine sediment delivery from bank erosion to prior estimates of total sediment delivery to the lake. He concluded that streambank erosion constitutes "25% of the average, annual fine-sediment load delivered to the lake from all sources." He also noted that the two largest contributors of fine sediment (the Upper Truckee River and Blackwood Creek) "account for slightly more than 80% of all fines emanating from streambanks." These findings suggest an obvious geographic focus for management attention, but they do not necessarily support the wholesale aggradation of SEZ's regardless of location in the watershed, or any equivalency in the importance or value of their restoration.

Sediment budgets for non-urban catchments in the Tahoe basin (e.g., Stubblefield et al. 2009) suggest the types of information that would be needed to discriminate "urban" from "non-urban" sediment sources, but such information is not readily available in a compiled form. However, even localized case studies such as Stubblefield et al. emphasize the variability of sediment sources, and they stress the importance of understanding the nature, intensity, and location of the dominant processes (see assumptions 2 and 3 above). As they note (Stubblefield et al. 2009, p. 164), "Previous work in the Tahoe basin suggests three dominant erosional processes are prevalent: landslides; stream bank erosion from both tributaries and the main channel; and gully, rill and sheetwash occurring on steep denuded areas in headwaters, i.e. badlands." Not one of these processes has a particularly "urban" fingerprint, and only the first is likely to be improved by SEZ restoration.

These complexities notwithstanding, an influential body of literature has long-emphasized the potential for urban development to induce high rates of soil delivery to receiving waters (e.g., Wolman and Schick 1967). However, the very high rates of development-induced erosion that have gained management attention (such as the figure reproduced below from Wolman and Schick, 1967) are almost invariably associated with the land-clearing stage of urbanization (the "total exposure" [of the land surface] in the graph below). Less commonly appreciated is the subsequent reduction in sediment yields from the land surface once the ground is covered, an intuitive but often overlooked outcome of paving over exposed soil and stabilizing infrastructure-threatening

landslides. In such settings, sediment loads that remain elevated above undisturbed levels are commonly due to the contribution of streambank erosion and incision from urban-increased runoff (e.g., Trimble 1997, Nelson and Booth 2002). Although this process is a consequence of increased impervious area, this is a useful indicator of such impacts only in the absence of effective stormwater management. Thus, avoiding or remediating such problems typically requires *runoff* control, not "erosion control." A lack of understanding the erosion-generating process(es) can therefore risk implementing mitigation that is not simply ineffectual but actually misguided.



- *Impervious cover*

Since Bailey's 1974 report, well over 4000 articles have been published in the scientific literature on urban development where "impervious" or "imperviousness" is mentioned in the title or abstract. Even the definition of the term has become more sophisticated, recognizing the critical differences between "connected" vs. "disconnected" (i.e., effective vs. ineffective) imperviousness that now underlies much of the recent technological and policy advances in stormwater management. There is no disagreement that impervious area is a key, landscape-scale indicator of urbanization and a host of associated disturbances, but any meaningful guidance for the 21st century should be using this metric, at most, as a starting point, not the sole criterion, for regulation and for tracking management success.

Degraded Stream Environment Zones (SEZ's)

As with imperviousness (as a broad indicator of "urban disturbance," and so surely worth minimizing), so SEZ's are valuable biotic and abiotic resources worthy of protection or restoration. Translating this recognized importance into effective and achievable management targets, however, requires something more than the blunt instrument offered by the guiding documents for the Threshold Evaluation. The discussion of the "Indicator status" beginning on page 5-13 of the 2015 draft exemplifies this critical shortcoming:

1. Preserve existing naturally functioning SEZ lands in their natural hydrologic condition:

Although asserted in the text to be "in attainment," I see no basis to make such a conclusion. It's apparently based on the presence of regulations for which "new coverage in SEZ lands has been prohibited unless it can be fully mitigated." I appreciate the faith placed in such regulations, but as a staff scientist who helped enforce stream and wetland protection for a decade in a less far-flung watershed than that of Lake Tahoe, the apparent lack of field verification leaves this discussion somewhat unconvincing. Are there any data to support this assertion?

2. Restore 25 percent of the SEZ lands that have been identified as disturbed, developed or subdivided:

The 1+ page effort to "explain" this indicator gives ample testimony to its unworkability, and thus the broad irrelevance of any such evaluation. The discussion is almost Talmudic in its effort to justify making an evaluation (*any* evaluation) for purposes of the Threshold Evaluation; the authors should simply eschew their assigned task of "evaluating status attainment" and abandon any pretense of coherence in the present guidelines, in order to serve the greater good of maintaining an overall tone of credibility for the remainder of the chapter.

That said, the goal of restoring riparian lands is a worthy one, and measures to track that progress are surely needed. Improving the current state of definition and evaluation would be a worthy task for the Threshold Evaluation. For example, is "disturbed" the same as "subdivided"? They are lumped together in this category, but presumably not *every* SEZ on subdivided property is been disturbed (and if it has been, they how can we have faith in the regulations touted in #1 above?). If an SEZ on subdivided land, presumed to be in the presently counted acreage of "disturbed, developed, or subdivided," is subsequently found to be in a functioning condition, does it now count as "restored" and so contributing to attainment of the 25% target?

3. Restore all disturbed SEZ lands in undeveloped, un-subdivided lands:

This discussion on p. 5-18 of the 2015 report raises the obvious question of why *all* SEZ's, regardless of location, are not targeted for ultimate restoration. As noted elsewhere (and as widely appreciated in the broader scientific literature), riparian zones are perhaps the single most critical, "sensitive" parts of a landscape. Why would such a goal not be articulated for the entirety of the stream and river network of the basin? Indeed, why would the restoration effort not be greatest where the impacts from surrounding land uses are most intense and problematic?

As noted above, in response to a peer review comment on the 2011 report—if the response to such comments is that the scope of this Threshold Evaluation is self-limited only to evaluating the criteria as previously administratively defined, then this is not a scientific peer review and the Threshold Report is not responsive to best available science—and it simply should not be represented as such.

4. Attain a 5 percent total increase in the area of naturally functioning SEZ lands:

Given the uncertainties noted above in terms of both definition of terms and on-the-ground conditions, it is difficult to see how any defensible evaluation of this threshold can be made. It is equally unclear why "5%" is a meaningful, credible value. If there are underlying economic, social, or political drivers that require a target that falls somewhat short of full restoration of *all* SEZ lands, then in the interest of producing a transparent, scientifically defensible document those drivers should be shared. Otherwise, this indicator stands in direct contradiction to the prior discussion of these critical landscape features.

SUMMARY AND RECOMMENDATIONS

Using a 40+ year old unreviewed report as the foundation for evaluation and management of soil-related resources in the Tahoe Basin is *not*, emphatically, "best available science." That said, the framework provided by Bailey (1974) is a credible one, and the revision that is so urgently needed here need not abandon all prior work

to constitute a credible replacement. Such an effort would likely follow steps similar to the following suggested sequence of activities:

1. Compile key pieces of watershed information and definition—topography (in particular, slope angle), geology and soils, hydrography (rivers, streams, wetlands). These are the "factors" of Bailey (1974).
2. Integrate these layers into relatively homogeneous landscape areas with respect to sediment-production susceptibility. These are equivalent to the "hazard classes" of Bailey (1974), although the manner of their integration requires greater transparency and objective criteria than is presented in that earlier document.
3. Evaluate the magnitude of soil erosion and delivery of each hazard class, distinguishing between natural production rates and those associated with land disturbance. This was accomplished in Bailey (1974) using somewhat intuitive and unquantified criteria, and without distinguishing "natural" from "disturbed" productivity in either relative or absolute terms. A "modern" approach would undoubtedly include some combination of prior (i.e., published) and new field measurements and observations.
4. Integrate the distribution and the magnitude of soil erosion into sediment budget for the watershed, at a level of accuracy and precision necessary to identify (1) major sources of sediment into the lake and (2) potential greatest increases in sediment delivery as a result of future development. This was done in only a rudimentary fashion in Bailey (1974), simply by identifying the "most susceptible" classes. Although a useful first step, this does not provide a basis for evaluating whether future development is likely to significantly impact sediment delivery to the lake. A good starting framework could be developed from texts such as Reid and Dunne's book, *Rapid Evaluation of Sediment Budgets* (1996, Catena Verlag GMBH, 164 pp.) or any of the 144 published works that have since cited it.
5. Articulate the goal of watershed management with respect to sediment management: is it no net increase in sediment delivery? A net improvement? Degradation of SEZ's that is no more than X%? Minimize degradation of SEZ's while accommodating a Y% increase in developed land? Any such criterion is nowhere apparent, which makes any exercise in setting disturbance limits or restoration targets completely arbitrary.
6. Develop a suite of alternative management approaches that achieve the articulated goal of #5, based on the understanding of sediment-erosion and sediment-delivery processes developed in #4. This might include some combination of outright impervious-area limitations (the only strategy advanced in Bailey 1974), stormwater control from either new development and/or existing development, erosion control, and/or instream channel stabilization.

A somewhat parallel exercise is also needed with respect to the protection and restoration of SEZ's. Without the sediment budgeting of #4 above, there is no way to know the degree to which degraded riparian zones are contributing to the sediment load of the lake, although recent studies (e.g., Simon 2008) suggest that these zones in general (even if not "degraded") may be a significant source basin-wide. However, "restoration" of a SEZ should not necessarily be synonymous with "channel stabilization," insofar as a naturally functioning river or stream may have significant, albeit localized, zones of bank erosion associated with channel migration. Channel erosion, of course, can be an entirely natural process that contributes to healthy stream ecosystems and the creation of critical habitats, and so invoking a simple numerical target for restoration (particularly if the implementation is simply bank-hardening streambank stabilization) is unlikely to achieve broader ecological goals that ultimately will support a healthy lake condition as well.

Even prior to such a comprehensive evaluation, not all "degraded SEZ's" have equivalent impact on sediment erosion and delivery, and not every locality is equal to every other with respect to net sediment contribution, broader opportunity for a broader range of benefits, or ease of achieving desired outcomes. Without some

framework for stratification, never mind an understanding of actual importance to the stated goals, there's really no point.

In summary, the present approach to evaluating the condition and the improvement in SEZ's is an overly blunt instrument with no apparent scientific basis beyond "more is better." The science has truly advanced in the last 40+ years; this Threshold Evaluation should, at minimum, acknowledge how much work remains to bring the protection of Lake Tahoe into the 21st century.

REFERENCES

Nelson, E J., and Booth, D.B., 2002, Sediment budget of a mixed-land use, urbanizing watershed: Journal of Hydrology, v. 264, pp. 51–68.

Rios, D.T., Chandra, S., Heyvaert, A.C., 2014, The importance of small urbanized watersheds to pollutant loading in a large oligotrophic subalpine lake of the western USA: Environmental Monitoring and Assessment, 186 (11), pp. 7893-7907.

Simon, Andrew, 2008. Fine-sediment loadings to Lake Tahoe. Journal of the American Water Resources Association 44(3), 618–639.

Stubblefield, A.P., Reuter, J.E., Goldman, C.R., 2009, Sediment budget for subalpine watersheds, Lake Tahoe, California, USA: Catena, 76 (3), pp. 163-172.

Trimble, S.W., 1997, Contribution of Stream Channel Erosion to Sediment Yield from an Urbanizing Watershed: Science, v. 278, pp. 1442-1444.

Wolman, M.G., and Schick, A., 1967, Effects of construction on fluvial sediment, urban and suburban areas of Maryland: Water Resources Research, v. 3, p. 451 464.

Review by Dr. Robert Burns

West Virginia University

Fundamental questions the reviewers should consider in their review:

Chapter 1. Introduction.

Does the Introduction Section provide sufficient background information necessary to understand the purpose and scope of the Threshold Evaluation Report?

Yes, the introduction adequately describes the history and process. The description of the various indicators used is sufficient. The description of how the various indicators are categorized and coded is important information. It is easily understood and provides a reader with a good sense of the report.

Are statements of fact presented in the Introduction Section sufficiently supported with appropriate references or original data and analysis?

I feel the introduction could have additional discussion about where the "reliable" information came from. The reader is left with needing to interpret what this means. I would suggest adding hyperlinks that take the reader back to previous studies and previous threshold findings.

Chapter 2. Methodology.

Are prescribed approaches for determining the status and trend of indicators relative to adopted standards clearly presented and appropriate?

Yes, these are clearly presented. This chapter is written in a manner that is perhaps clearer than the Introduction chapter. The use of definitions, arrows, charts and colors makes sense and assists the reader in understanding.

Is the prescribed approach for determining the level of confidence in status and trend determinations clearly presented and appropriate?

Yes, no issue here at all.

Are the approaches prescribed to determine an interim target and attainment date for an indicator reasonable, given data and funding limitations?

I find it curious that the methods section suggests specific tests (for example linear regression). All methods of statistical analysis should be considered and used when appropriate. These issues are complex and very inter-related across several variables. Perhaps the use of newer, more detailed statistical analyses would be appropriate. An example might be use of structural equation modeling to explain these complex relationships. Discussing just 1-2 statistical analyses in the methods section limits your ability to be progressive.

Do the authors clearly articulate the sources of uncertainty in the analysis?

Yes, this is done appropriately.

• **Indicator and Indicator Reporting Category Evaluations** – Is the write-up associated with each indicator evaluation clear and complete? Are the analytical methods appropriately applied in the determination of an indicator's status, trend, and confidence? Are there other or different analyses you would recommend to

evaluate indicator status or trend? Are statements of fact or conclusions supported by appropriate references or original data and analysis?

Chapter 9. Scenic Resources.

This chapter is an inventory and evaluation of the myriad of various scenic resources. The list is certainly exhaustive, and rates a majority of all scenic resources as "at or somewhat better." This seems to be a very high proportion of scenic resources that are very good. If the management and stakeholders agree, than this method may be appropriate.

The chapter is well written and uses a logical explanation process. However, I would suggest that the scenic resources should be evaluated by users of the area in addition to an inventory with manager-developed perceptions. If user surveys have been conducted, that information would support managers' perceptions and would be another measure of quality assurance.

• **Recommendations** - Are recommendations appropriately supported by report findings and conclusions? Are proposed recommendations related to policy and on-the-ground project implementation supported by best available science? Are remaining uncertainties clearly articulated?

As noted previously, the use of stakeholder, community and user group perceptions for Scenic Quality would be helpful for the managers to know that their opinions do not vary (or do vary) from those of the stakeholders. I would suggest the use of stakeholder data.

Chapter 11. Recreation.

The recreation survey section is rudimentary and, in reality, a loosely cobbled together set of data. As it exists I doubt the databases are valuable in management decision-making. The data should be analyzed more thoroughly and presented in the report in a way that is useful. The data should be segmented across various settings within the management area, and sociodemographic data should be analyzed as well.

I have concern that the data were collected by various different groups, and not by one agency. While this method can be perfectly fine if it is managed properly, the methods used should be outlined in the Methods chapter. The USFS NVUM process is referenced, but then quickly dismissed. There have been 3-4 rounds of data collected at the Lake Tahoe Management Unit. While the 2015 data may not yet be available, it would be useful to see the previous data analyzed and shown, at least for the high use sites.

The PAOT data should be analyzed for this report, not simply referred to as being historically bad (page 13). This report should show trend data over time, and include the same types of graphs and arrows as seen in the methods section.

I am not confident that the satisfaction questions will answer what is needed for progressive park management. For example, satisfaction with recreation access and safety are lumped into one category. A quick look at previous research would show that these should not be grouped together, as they are very different questions. The lowest score is seen for this item. But is the problem with safety or access? And where does the problem exist? The data are presented at the end of this report in a confusing manner. I am concerned about the very low ratings of satisfaction for many of these items.

Review by Dr. Stephanie Guildford

University of Minnesota, Duluth

Review of the Tahoe Regional Planning Agency's 2015 Threshold Evaluation Report

Thank you very much for the opportunity to review this report. As a limnologist who has been concerned about eutrophication in large lakes around the world, I welcome the chance to learn more about iconic Lake Tahoe. The TRPA is to be commended for its extremely comprehensive work with such a large and varied number of groups of stakeholders. It is clear that in the overall sense Lake Tahoe is well protected and showing positive signs that the efforts to protect it and remediate it are working. I reviewed chapters 1, 2 and 4, the Water Quality chapter. My comments specific to each chapter are below.

Comments on Chapter 1 Introduction

History

The background and historical information is adequate and appropriately brief; however, it could be even briefer and more specific if it contained a simple table or diagram of the timeline of events impacting the Lake Tahoe ecosystem. This section should include a few statements addressing important environmental challenges faced by the TRPA that are imposed beyond the watershed boundary including climate change, atmospheric transport of nutrients and pollutants and invasive species.

2015 Threshold Evaluation Report

This introduction to this section acts as a reasonable guide to how to navigate the individual chapters for each of the nine threshold categories. I found the reference to Figure 1-2 confusing at the point when the reader is referred to "As can be seen in the right hand column…." – this is confusing and does not serve to illustrate the statement about types of standards.

Description of Indicator Summaries

This section is in general straightforward; however, it becomes confusing at the heading "Implementation and effectiveness". It seems this subsection should be more logically entitled "Progress toward attainment of threshold standards".

Evaluation of Management Standards and Policy Statements

This section of the Introduction was confusing to me as it did not seem relevant to the Water Quality Chapter. Although some Standard Types listed as Management and Policy are listed in Table 4-1, few of these standard types had related indicator summaries in Chapter 4. This section of the introduction should be clarified as to what section or sections of the report these statements are applicable to.

Comments on Chapter 2 Methodology

The introductory paragraph for this chapter is a bit misleading. Item 2 in the list on page 2-1 indicates this chapter includes an outline of the indicator summaries. This information is actually provided in Chapter 1.

Evaluation of numerical standards and management standards with numeric targets

Determination of indicator status

This section is problematic as it provides an example of how it may be inappropriate to express target attainment status as a percent calculation followed by a detailed example for Secchi disk depth which was the example previously provided. Light is absorbed and scattered logarithmically with depth and this makes it difficult to compare Secchi disk depths directly to variables that change in a linear fashion. In general the Secchi disk depth should be logged to perform calculations. It would be best to provide an example using a different type of indicator such as TP.

Evaluation of indicator trend

The use of a simple linear regression on non transformed data may not be providing appropriate trends. As discussed above calculations involving Secchi disk measurements should be transformed. Other data, where the ranges of values are large, should also be transformed to avoid possible sensitivity to a few high or low numbers. Many researchers use the Sens slope rather than regression for this type of data.

Confidence in status and trend determination

It is good to see that you provide assessment for the different components of the indicator assessment. While a measurement in itself might be very robust, the analyses of temporal and spatial data is much less likely to be statistically robust. This is a good reason to have more than one indicator to address a particular issue.

Evaluation of management standards and policy statements

This section of Chapter 2 was not really relevant to the water quality chapter; however, I did notice that this section appears to be a repeat of what was previously provided in Chapter 1 (the Introduction) on page 1-10.

Sources of status and trend data information

This information appears in the indicator summaries and could possibly be eliminated from this chapter.

Overall assessment of Chapters 1 and 2

Adequate. Some duplication. It might flow more readily if the Introduction was briefer. The section on Description of Indicator summaries might fit more logically into Chapter 2 which I would entitle Guide to Indicator Summaries rather than methodology.

Comments on Chapter 4 Water Quality

General issues

1. The use of annual averages to analyze trends is problematic as many variables are driven by a few episodic events
2. Many references were missing or if cited not readily available.
3. Where does the primary productivity go?
4. Modelling efforts are needed to link these separate data sets in a meaningful way to water quality. For example, the Sparrow model or more detailed SWAT model and others used in the Great Lakes could be helpful

Specific Comments

Introduction

Page 1, 2nd paragraph lists the six threshold standard categories. This section requires clarification in two ways. The categories should be more clearly defined. Depth criteria for pelagic and littoral water should be included. Tributaries are generally considered to be the conduit of surface runoff so this leads to confusion when there is a category called "surface runoffs." This category as reported on here is more accurately defined as storm water outflow. The ground water and "other lake" category are not defined and although they are listed in Table 4-1 they do not appear again in the report. This should be changed or explained.

Several references here and throughout the chapter (and in chapters 1 and 2) were either not included in the list of references or not available online. This leads to the reader wasting time looking for citations and becoming frustrated.

In the next TRPA Threshold evaluation report I strongly recommend including metric units wherever imperial units are used throughout the report. Although the public is more familiar with imperial units, scientists are used to thinking in metric units and it is distracting to have to convert from imperial to metric units.

Page 1 last paragraph, it would be helpful to include the date when the value statements were developed to provide context. If these are the value statements coined in 1982 they could be understood in that context. Are these value statements examined periodically to determine if they still reflect the views of the TRPA.

Fine sediment should be quantitatively defined.

Page 2 paragraph 3. It would be informative to provide specific examples of some of the invasive species and changed nearshore conditions that are alluded to in this paragraph.

Lake Tahoe's pelagic waters

The designation of greater than 10 m is too shallow. Other studies consider the littoral to be less than 20 m. The scientific definition for the littoral is the 1% light level. In Lake Tahoe this depth would be much deeper than 10 m. By classifying water that is deeper than 10 m but receiving adequate light to drive benthic algal photosynthesis important information pertaining to changes in the nearshore and littoral may be missed.

The number of indicators for the pelagic is surprisingly small and the fact that there are no indicators in common with the littoral, tributary and stormwater runoff data is unfortunate. It makes it difficult to track and understand the linkages between the land and the lake and, therefore, demonstrate the impact of changes on the land to the lake. I recommend the inclusion of TSS, FSP, TN, and TP for the Pelagic, Littoral, Tributary and Stormwater runoff indicator categories. I also recommend adding phytoplankton chlorophyll a as a measure of phytoplankton biomass in both the pelagic and littoral. In the pelagic the addition of particulate stoichiometry (particulate carbon, nitrogen and phosphorus) would provide much needed information about the nutrient status of the pelagic phytoplankton. It is possible that some of these measurements are already being made and could easily be included as indicators and importantly these would be "linking indicators" for the Threshold Evaluation Report.

There is inconsistency between information in Table 4-1 and the indicator summaries. Nitrogen Loading is listed in the Pelagic indicator category as a numerical standard, but there is no indicator summary. Pollutant loading is listed as a management standard type for N, P and Fe but no summaries were provided. A standard listed as "Record of threshold standard exceedance" is confusing with no explanation in Table 4-1 or in the indicator summaries. These inconsistences should be explained.

Pelagic Lake Tahoe Annual Average Secchi depth

Human and environmental drivers

This is a good description of the factors that contribute to the Secchi disk depth value and of the anthropogenic and environmental drivers that affect the Secchi disk depth. A driver not considered here but important to the

Secchi dynamics is climate change. Climate change may have overarching indirect impacts on the components of the Secchi disk depth and specific impacts such as drought creating dust, increased temperature modifying the depth of stratification and therefore the light environment for phytoplankton in and below the epilimnion.

Secchi trends

This discussion raises several questions. Where are the separate winter and summer data? What are the dates of winter data? Has the timing of the winter and summer data changed since 1967 as a result of climate change. Meteorological data might be helpful in understanding trends especially more recent trends. The statement about a decline in small algal cells needs documentation. The same statement is made in the State of the Lake report for 2015 with no other information or reference. It is necessary to explain how this could contribute to a change in Secchi depth, but not a change in total chlorophyll a concentration (as presented in the State of the Lake Report (SOL)). Similarly it is necessary to explain how the shallower mixing depth contributes to a decline in Secchi.

Secchi confidence

I agree the Secchi disk depth is a very useful indicator and should be continued. It is as you point out impacted by several factors so it is important to convey that in any presentation of the data and in general avoid making much of year to year changes.

Programs and actions

TMDL reports. The 2014 handbook was not accessible online. The 2010 publication specifies phosphorus and fine sediment particles need to be reduced as well as nitrogen. It is important to be specific as different approaches and costs are associated with these three different pollutants.

Recommendations

Nutrient reduction as recommended in the Sahoo et al 2015 publication is an important recommendation. As above there should also be an effort to determine how the reduction of N and P using different approaches differ in cost and effectiveness.

Other comments

Fine suspended sediment is not included as an indicator in offshore yet this is what makes the major contribution to scattering of light (Swift et al 2006). Swift demonstrated that inorganic particle scattering primarily due to the < 4 µm size inorganic particles is the major contributor to Secchi disk measurements in pelagic Lake Tahoe. Based on the Secchi disk measures and the chlorophyll concentrations available in the SOL 2015 report, one would have to conclude that the decrease in Secchi disk transparency since 1984 must be due to an increase in suspended sediment as chlorophyll has remained essentially constant. Clearly the measurement and reporting of fine suspended particles in the pelagic would be relevant in the TER. To be of the most use, I suggest reporting the changes in FSP < 4.0 µm in the tributaries, littoral and pelagic in order to make the connection between the source of particles and impact on the lakes clarity. As Swift et al (2006) state these data are needed to run a model that will provide a linkage between erosion control measures and the lake's response.

The designated State standard of 29.7 m is based on the mean Secchi depth from 1967 to 1971 while the designated State standard for the Vertical Light Extinction coefficient (VEC) is not to exceed 0.08 (m^{-1}). The state standard for VEC has already been achieved several times since 1985. There is a disconnect between these two standards. A general equation to predict VEC from Secchi depth in waters not highly influenced by colour is:

VEC=1.7/Secchi Depth

Using this equation to estimate VEC for Lake Tahoe would generate a VEC of 0. 057 (m^{-1}) to meet the Secchi standard. This is almost exactly the value for VEC measurements reported for Lake Tahoe in the years 1971 to 1977.

It seems that there should be a more similar standard (or at least a clear explanation for each standard) to avoid confusion and doubt by the general reader who sees the goal for VEC being met but the goal for Secchi depth seeming a long way off.

Pelagic Lake Tahoe Phytoplankton Primary Productivity

Relevance

The annual average areal rates of carbon fixation reported for Lake Tahoe have been increasing in a linear fashion since measurements began in 1967. This is an interesting and important trend for Lake Tahoe. Primary productivity is a very complex indicator, e.g. it integrates light availability, light absorbance by phytoplankton as well as nutrient availability and modification due to consumption of the carbon fixed, so that the trends observed in Lake Tahoe could be the result of several factors impacting PP individually or in combination. Without understanding the underlying factors it may be inappropriate to conclude that the increasing rates of PP necessarily represent a negative trend for the health of Lake Tahoe. Assuming the methodology has remained completely unchanged, the trend should be explainable by understanding the components of primary production. The biomass and photosynthetic efficiency of the organisms at a given irradiance comprise the areal rate of carbon uptake at any given time and location in the lake. Thus any changes in algal biomass, pigment composition, photosynthetic efficiency, and light quality and or quantity could result in a change in areal carbon fixation. It is difficult to know which if any of these factors have changed over the course of PP measurements in Lake Tahoe because although carbon uptake is measured routinely at several depths it is not clear if there are measurements of chlorophyll and photosynthetic efficiency at the same depths. These data would be needed along with light data to understand what changing factor is driving the increase in PP. Data found in the primary literature report that there have been changes in the phytoplankton community composition since monitoring began including a shift to smaller centric diatoms which is similar to trends in other large lakes undergoing re-oligotrophication (Reavie et al. 2014, JGLR 40:3 pp 618-639). Unfortunately there is not much information on the pico phytoplankton in Lake Tahoe. These tiny algae may be becoming more important in Lake Tahoe and to understand if smaller diatoms and smaller algae in general may be the reason primary productivity has increased without a concomitant increase in chlorophyll a would require information about the photosynthetic efficiency of these smaller species. The fact that chlorophyll a is not increasing in the pelagic waters of Lake Tahoe suggests that the measure of PP in the pelagic is not a strong indicator of lake health without additional information on the components of PP and importantly on the fate of PP. The lack of change in chlorophyll a concentration could also be explained by changes in loss rates of PP to grazers and or sedimentation; however I was unable to locate any recent studies exploring these processes in Lake Tahoe.

Human and Environmental Drivers

Goldman et al (1993) demonstrated using enrichment bioassays that historically Lake Tahoe phytoplankton were nitrogen limited. However, beginning in 1980 phytoplankton were much more likely to be limited by P. The primary literature and the grey literature on Lake Tahoe consistently state that N and P are limiting phytoplankton growth in Lake Tahoe. If phosphorus is currently the limiting nutrient then the focus should be on P reduction rather than both P and N. Enrichment bioassays can be quite useful and they are cost effective; but, they are not always a reliable indicator of the processes occurring in situ. Phytoplankton incubated in containers during enrichment bioassays are cut off from nutrients potentially obtained from atmospheric gases and from regeneration by grazers and as a result these enrichment experiment results could provide an incorrect assessment of the limiting nutrient in situ. There are several simple physiological and composition measurements that could be made on water from the epilimnion of the pelagic of Lake Tahoe to document the

nutrient limiting phytoplankton growth at any given time of the year (Guildford and Hecky 2000 L&O 45: 1213-1245).

The majority of P entering Lake Tahoe is most likely originating in the basin and thus can be much more readily controlled than N which can have a major atmospheric component. The Laurentian Great Lakes recovered from eutrophication primarily as a result of the Great Lakes Water Quality Agreement implemented in 1972 and most recently revised in 2012. This agreement sets phosphorus loading targets and limits on point source emission and detergent use that have been demonstrated to be effective in reducing algal biomass. The Great Lakes recovery occurred even as nitrogen concentrations in the lakes continued to rise as a result of agriculture and transportation emissions. This recovery was driven by regulating point sources of P. Unfortunately, the low hanging fruit of P control has been picked and the managers of the Great Lakes now face the daunting task of trying to control non-point sources P from intensive agricultural areas particularly in the lands riparian to the western basin on Lake Erie. Documented increases in soluble reactive phosphorus concentrations in combination with warming temperatures and more intense precipitation events are driving potentially toxic blooms of the cyanobacteria *Microcystis* (Michalak et al 2013 PNAS 110: 6448-6452; Scavia et al. JGLR 40: 226-246). The P cycle in the Great Lakes has also been dramatically impacted by invader zebra and quagga mussels (see section on AIS below for more comments on AIS in Lake Tahoe).

Confidence in PP as an indicator. (High). I agree that you can be confident in the trend; however, as stated above the complexity in PP makes it difficult to interpret and therefore to be a useful indicator.

It is telling that there are no applicable State and Federal standards for phytoplankton primary productivity. This is not a commonly used indicator for eutrophication. Chlorophyll a is more commonly measured and although problematic in some ways it is a much more direct measure of phytoplankton biomass, and related problems such as algal blooms and potential oxygen demand through decomposition, than primary productivity.

Pelagic Lake Tahoe Vertical Extinction Co-efficient

I think that Vertical Extinction measurements should be continued. They provide a complimentary measurement to the Secchi with a different sensitivity to particulate matter, and can be quantitative and analyzed to provide more insight than Secchi alone. VEC can be used to calculate the depth of the euphotic zone, the mean water column light intensity both measures are useful in understanding and quantifying primary productivity. However see Section on Secchi depth above for concern regarding the State standard.

Littoral Lake Tahoe

As stated previously the criteria for littoral (< 10 m) may mean important information is unavailable about the health of the nearshore areas of Lake Tahoe. I appreciate from the information in the introduction to this section that recognition of the importance of the nearshore in Lake Tahoe is a relatively recent phenomenon as it has been in many economically and environmentally important large lakes and that it will take time to develop an appropriate monitoring program and generate data to evaluate trends. For example, the Great Lakes Water Quality Agreement (revised in 2012 specifically calls for development of a Nearshore Framework for monitoring and managing the nearshore environment. I recommend including common indicators in the littoral that would permit LTRPA to assess the impact of land use changes in the watershed on littoral and pelagic waters. Vertical light extinction, TN, TP, phytoplankton chlorophyll a, TSS and FSP measurements in nearshore waters taken in waters with a depth of about 10 m would be appropriate as, at this depth, the effect of episodic resuspension would be minimized.

As noted above for the pelagic waters indicator summaries, it was difficult to reconcile the information from Table 4-1 with the indicator summaries.

Nearshore water clarity

In Table 4-1 the indicator standard "Sediment Loading" is listed with units of NTU. In Table 4-2 the summary of status and trends lists the indicator "Turbidity" at and away from stream mouths. In the indicator summary the indicator is referred to as "Nearshore water clarity" While all these terms are related they are not at all equal. It is necessary to be more clear and consistent about what is being measured and assessed and reported on in the different tables and summaries. The detailed discussion in the Recommendations section for this indicator summary implies that this indicator standard is being re-evaluated with the idea of using beam transmission as an indicator or water clarity rather than NTU. I strongly support this recommendation.

There is some confusion in understanding the data presented in this section on littoral water clarity. After reading and rereading I believe I now understand that the data presented as box plots in the first figure in this section are data collected at the 25 m depth and the data presented in Figure 1 on page 4-37-38 are from a pilot study at the 7 m depth. Data at the 7m depth would be expected to be more representative of the nearshore than the 25 m depth. Although the indicator summary Table 4-2 provides status and trends for turbidity at stream mouths and away from stream mouths, there is no information or data in the indicator summary about the spatial variability relative to stream mouths in either the long term data collected at the 25 m contour or in the pilot study done at the 7 m contour.

I recommend VEC measurements be included as a littoral indicator for nearshore waters that are deeper than 7 m but shallower than 20 m. These measurements would provide a linkage to the VEC measurements in the pelagic and would also be useful for modelling periphyton growth in the nearshore.

There is a real need for a nearshore water clarity indicator that can be compared to tributary load data to explore patterns and determine factors that reduce transparency along the gradient from nearshore to the pelagic.

Nearshore attached algae

It is good to see that the TRPA has adopted attached algae as an indicator of littoral health for Lake Tahoe. The accepted light level cut off for when respiration exceeds photosynthesis is 1% and in Lake Tahoe, attached algae would be expected to exhibit positive growth at depths of 20 m and even deeper. It is surprising that the monitoring for attached algae in Lake Tahoe is done at 0.5 m. At this shallow depth attached algal growth and biomass accumulation is expected to be significantly impacted by physical processes such as light inhibition of photosynthesis, high ultraviolet light exposures and by wave action. I recommend that the standard for attached algae be re-evaluated.

There is confusion in trying to reconcile Table 4-1, 4-2 and the indicator summary for attached algae.

The table of data for attached algae in the indicator summary does not provide units or an explanation for the > and < 6225 feet or 1000 days. It is very confusing. Reporting a mean value for the whole lake sampling is not informative for this indicator as it is so heterogeneous in space and time. The literature referred to were not readily available. Attached algae is an important indicator however based on this section of the report it is not clear that the monitoring or analysis are adequate to provide useful information.

Nearshore Aquatic invasive species

It is good to see that Lake Tahoe has a boat inspection program as this must be one of the main vectors for introduced species. Although it is good to see a large overall effort to protect against AIS it might be helpful to be more specific and to identify and prioritize based on type of organism and potential threat based on what has been learned in other water bodies. In the Laurentian Great Lakes increased water clarity as a result of filtering by the aquatic invader dreissenid mussels has resulted in dramatic increases in attached algae (Higgins et al. 2005, JGLR 31:547-563).

Tributary water quality

The indicators in the tributaries are the most thoroughly covered indicators in the report however the use of the annual averages even when plotted as loads does not provide important information about the extreme and episodic events that would be expected to be the most important factors influencing long term trends and nutrient and sediment delivery into the lake.

Tributaries suspended sediment concentration

The data are difficult to assess because of the problem of sample number and timing. There is a good discussion of this problem, and it appears that there has already been an effort to address this problem. I agree that continuous and or automated monitoring for some parameters in some streams would be valuable and provide a more realistic representation of the problem and the trends. The programs and actions to remediate suspended sediment in tributaries was vague for the two most problematic tributaries Blackwood and Ward. It would be good to see what actions if any can be suggested to remediate these undeveloped areas.

Tributaries total phosphorus concentration

The graphs for five of the seven tributaries were missing. Similar to the suspended sediment data the TP data in tributaries are difficult to assess due to temporal variability in both number of samples and actual concentrations. The data generated from continuous and automated monitoring of some variables will provide a more realistic assessment of improving or worsening trends for TP. These continuous monitoring records should help target problem locations and times and aid in understanding what mitigation actions are likely to be helpful and possible to implement.

Tributaries Total nitrogen concentration

As for TP and TSS the use of annual means is problematic when it is clear that inputs are extremely sensitive to weather events. The continuous monitoring approach is clearly needed. It is evident that efforts are ongoing to reduce human impacts and this is to be commended, but the future monitoring should help these efforts to be targeted more efficiently. It should be noted that in the Great Lakes TP inputs have been reduced resulting in reduced chlorophyll while at the same time TN has increased due to anthropogenic activities. While high concentrations of nitrogen can impact water quality in terms of algal composition and at extremely high concentrations threaten fish and even humans, there is not evidence that over the course of monitoring of the Great Lakes that nitrogen inputs have resulted in elevated chlorophyll. The Lake Tahoe report frequently states the need to control both N and P to control algal productivity and biomass. This may be the general consensus in the Lake Tahoe literature but it is not in the wider freshwater literature. Economic costs of controlling both may impact the ability to address specific causes and mechanisms to improve and maintain water quality.

Tributaries suspended sediment load

This calculation is helpful and is useful to evaluate large spatial and temporal scale changes. It will be much more powerful when continuous monitoring of turbidity is available in addition to stream flow. It is good to see that based on this data there has been considerable reduction in erosion.

Tributaries fine sediment load

These data show no trend although measurements have been made since 2002. It may be a problem of reporting one size range (< 40 µm) rather than specific bin sizes representing different classically defined particles such as silt, clay and colloidal size. I determined from the literature I was able to access that these different sizes are likely being measured as part of the FSP measurement and could be reported.

Total phosphorus load and Total nitrogen load

These data are a helpful way to synthesize the TP and TN concentration data and gage the impact of these changing TP or TN concentrations to the lake. These graphs demonstrate the importance of climate to the loading to the lake.

Surface Runoff

I agree that controlling storm water runoff from urban areas is an important part of protecting water bodies. As it stands this section is not entirely understandable. It is odd that there are no P, N or suspended sediment concentration data presented for this section, yet there are plots of TP, TN and FSP loads to the lake. There should be an explanation about why the concentrations are not shown. The loadings reported are a very small proportion of the loadings resulting from the tributaries. It would be helpful to develop a quantitative model to help understand how reductions in stormwater runoff nutrient concentrations would be expected to impact the lake.

The units for FSP are reported in weight units but the FSP loading data for the tributaries is reported as numbers of particles. This makes it impossible to compare the relative contributions to the lake.

Overall comments and conclusions

The water quality of Lake Tahoe is being evaluated with an impressive number of indicators and the record of data for many of the indicators is long and robust. This information is very valuable in assessing trends and modelling future conditions. I can understand the challenge of synthesizing this large data base and distilling it down to a 5 year evaluation report. Although the report provides detailed summaries of many indicator standards for the various components (indicator categories) of the lake, the lack of consistent indicator standards across indicator categories made it a challenge to understand how changes in the watershed impact the pelagic, littoral, tributary and stormwater runoff. For example, while it is clear that changes in TP loading from tributaries would be expected to impact littoral and offshore waters with respect to phytoplankton biomass, transparency and primary productivity, it seems that reporting TP concentrations in the littoral and the offshore in addition to the tributaries and stormwater would provide the reader with a more tangible and transparent way to assess improvements to erosion in the watershed. The flow of the report would benefit from a diagram to map the indicators and linkages in the watershed, littoral and pelagic of Lake Tahoe. Another challenge of presenting such a long and detailed record of historical data is that extreme events are often masked by reporting annual averages. Trends exhibited by plotting extreme events rather than annual averages could reveal important changes otherwise missed.

Review by Mr. Gary Hunt

TRC Environmental Corps

Tahoe Regional Planning Agency (TRPA)-Threshold Evaluation Report

June 30, 2016

Introduction-Chapter 1

Sufficient information is presented to allow for the reader to understand the purpose and objectives of the Threshold Evaluation. Data on the types and numbers of businesses by industrial sector would have been especially valuable as related to the air quality portion of the peer review. It would have been helpful also knowing which of these industries (and how many) had permits related to discharges into the Lake Tahoe environment. For example, water discharge permits, air emissions permits etc. If not provided in this section then it is recommended that this information be provided in the media specific sections of the report as appropriate. It would be helpful if the total number of primary and secondary residences located in the study region were also listed. As related to the air quality portion of the Threshold Evaluation it is also recommended that the numbers of wood stoves in place in residential and commercial properties and the numbers of vehicles owned and operated by these residents be also provided. Again this data can be provided in the Introduction section or the Air Quality section (Chapter 3) of the report.

Some portions of the Introduction chapter (commencing with the section entitled threshold standards) actually relate more to methodology employed in preparation of the 2015 TRPA Report. This information on pages 1-5 to 1-11 is somewhat redundant to information currently contained in Chapter 2 entitled Methodology. It is recommended that this material currently presented in Chapter 1 Introduction be incorporated into Chapter 2 Methodology.

Methodology-Chapter 2

Methodology employed was well described and clearly presented. Often times example calculations were provided. This was true for the Determination of Indicator Status and Evaluation of Indicator Trend sections. This chapter reflects revisions made by the report authors in response to the peer review of the 2011 TRPA report. For example, status scores were used in calculation of indicator status even when available data were insufficient. Confidence scores, as well as, indicator status and trend scores, however, were not artificially biased high due to insufficient data or lack of standards as was the practice in preparation of the 2011 TRPA report.

(Note: Readers should view only color versions of the report (hard copy or electronic on screen) to fully appreciate report contents. The methodology used and the actual results for each threshold category are hard to follow when viewed in black and white).

Noise-Chapter 10

The noise chapter represents an improvement to what was presented in the 2011 TRPA report. Responses to comments on the 2011 TRPA report offered by this peer reviewer are reflected in the 2015 report. Based upon the 2011 peer review "many of the noise monitoring thresholds, especially those related to single noise events

were not analyzed for this (2015) evaluation and therefore received a status of unknown". In addition the TRPA instituted a more rigorous monitoring program in the majority of the noise areas commencing in 2011 and continuing through calendar year 2015. This included monitoring at more sites for at least seven (7) days during the critical calendar period of May 15 to October 1. This reviewer endorses this practice over time as these data will allow for a more reliable evaluation of indicator status versus thresholds, data trends, confidence and threshold attainment status.

General Comments and Recommendations

1] The noise program remains too complex and resource intensive. There are too many indicators, land use categories and numerical thresholds that need to be monitored to evaluate data trends and attainment status.

2] It appears that current programs and corrective actions are not effective in reducing noise levels on those highways with a 55 dBA standard. Review and evaluation of noise standard values is recommended. For example, if trends indicate future attainment is not likely numerical standards may need to be increased. In the case of highways standards range from 55 dBA to 65 dBA for cumulative noise events (CNEL). Attainment with the 55 dBA standard has been a challenge while attainment with the 60 dBA and 65 dBA standards has been less of a challenge. Consider adopting a 60 dBA standard for all highways.

3] The TRPA hired a noise consultant in 2011. As a result dramatic improvements were made to the noise monitoring program. These included but were not limited to the following: 1) instrument calibration. 2) increased frequency of monitoring (# days and # sites). These measures, also recommended by this peer reviewer in 2011, resulted in more reliable and representative noise measurements across the study region.

4] Recommendations for additional actions, findings and conclusions are frequently contained within the individual data evaluation and interpretation reports for each indicator category. These need to be consolidated to remove redundancy and should appear as appropriate in the report summary and conclusions chapter.

Comments and Recommendations- Single Noise Event Categories

1] Threshold standards for the majority of the single noise event categories could not be evaluated due to insufficient field data. A severe lack of data does not allow for adequate evaluation of many of the single noise event indicators. As a result, the effectiveness of existing regulations cannot be characterized. There are, again, too many single noise event standards. These need to be reduced or consolidated in order to avoid insufficient data determinations in future years.

2] The current criteria of "zero exceedances of the threshold standards" for a single noise event is unrealistic. Attainment in the future seems unlikely, as a result. An alternative threshold attainment measure should be sought. Non-attainment should not be based upon a single exceedance of a standard but rather make use of all data collected. The data set should include all compliant measurements, as well as, all exceedance data. In this manner attainment can be defined as a % value (total compliance values/total # measurements). Statistical analyses of data for each category should also be applied. This approach is especially important to those noise standards characterized as single noise events (eg motorized water craft).

Comments and Recommendations- Cumulative Noise Event Level (CNEL) Categories

1] The CNEL data base is much improved since 2011 when the enhanced monitoring program began. Indicator status, data trends, confidence determinations, threshold attainment, and overall program effectiveness can now be evaluated with the availability of more reliable and representative data sets for many of the CNEL noise categories.

2] This reviewer agrees with the current practice of establishing data trends by using annual mean CNEL values. Attainment status should also be based on the same annual mean CNEL values representing all monitoring

locations within a given land use category. The current practice of using the maximum 24 hour CNEL value to determine indicator and attainment status is overly conservative.

3] This reviewer endorses the following statement found in the recommendations section (Modification of the Threshold Standards or Indicator Section) for many of the CNEL land use categories in the noise chapter. (pgs. 10-17, 10-20, 10-26, 10-32, 10-35, 10-38, 10-41)

"O*utside Peer review for both the 2011 and 2015 Noise Threshold Evaluations have suggested that the attainment of CNEL should be based on the percentage of events that exceed the threshold rather than being based on a single exceedance and that TRPA should report on the number of locations that exceed the CNEL standard rather then the magnitude of the exceedance (TRPA 2012; Kerr et al. 2016)."*

4] Statements made regarding the effectiveness of the noise control program, threshold attainment and attainment dates are not supported by data presented for four of six of the highway categories. These statements are found in the Implementation and Effectiveness sections for these highways (pgs. 10-41, 10-47, 10-56 and 10-59). Revisions are warranted.

5] There are too many indicator categories for CNEL noise. Consolidation/combination of existing land use categories should be considered if this can be justified. There are currently five (5) separate numerical values used as noise standards (See Table page 10-3). Can this number be reduced to three (3)? (For example 45, 55 and 65 dBA).

Air Quality-Chapter 3

General Comments

1] In general the majority of the air quality indicators were found to be in attainment with the associated standards or indicator threshold values. Trends in most cases indicated moderate improvement. Air quality in the Lake Tahoe Region is well defined and the measures in place to improve air quality moving forward are working.

2] Recommendations for additional actions, findings and conclusions are frequently found within the individual data evaluation and interpretation reports for each indicator category. These should be consolidated to remove redundancy and placed as needed in the summary and conclusions section of the report. (Note: This reviewer was not assigned to peer review the report summary and conclusions section).

3] Only emissions from California vehicles are accounted for in the air quality indicator categories. CARB emissions estimates are used, for example, to develop daily and annual emissions for NOx. These data, in turn, have been used historically to monitor attainment with the NOx threshold standard. It appears that emissions from vehicles registered in the state of Nevada that operate within the Lake Tahoe region are not accounted for in any of these emissions estimates. Further, if Nevada vehicle emissions standards are not equivalent to California vehicle emission standards then continuous improvement in air quality may not be achievable. It is recommended that the revised TRPA report address the impacts associated with tailpipe emissions from vehicles registered in Nevada. For example, should Nevada consider adopting California vehicle emissions standards?

Carbon Monoxide (CO)

1] CO data indicate that concentrations for this indicator are well below the strictest applicable standards. This is true for both the 1-hour and 8-hour CO standards. TRPA has concluded that the overall status is considerably better than the target, that the trend shows moderate improvement with a moderate degree of confidence. This reviewer agrees with this assessment.

2] No interim target or target attainment date is offered as the Lake Tahoe Region is currently in attainment with the strictest standard. TRPA has also concluded that current regulatory programs, policies and actions directed at reductions in carbon monoxide emissions especially from motor vehicles have been effective. Existing programs and actions will remain in place for continued control of carbon monoxide emissions. This reviewer agrees with these findings and the recommended course of action.

Ozone- (O_3)

1] TRPA maintains a 1-hour standard for ozone of 0.08 ppm (v/v). This standard is more stringent than the corresponding standards for both California (0.09 ppm) and Nevada (0.10 ppm). TRPA recommends that the current standard of 0.08 ppm remain in place. This reviewer agrees with this recommendation.

2] Ozone data presented for the 3-year average of the 4th highest 8-hour concentration indicates that the Lake Tahoe Region has never violated the historical Federal Standard of 0.080 ppm or the revised standard of 0.075 ppm adopted in 2008. Despite conditions that TRPA defines as "at or somewhat better than target" the long term trend shows moderate improvement over the calendar period 1986,when monitoring began, until 2014, the most recent year where monitoring data are available. The data and statistical analyses show a decrease of 0.53% (or 0.0004ppm) per year. This peer reviewer is in agreement with these findings.

3] Ozone data presented for the 1-hour average supports an indicator or threshold status of "at or somewhat better than target" and a long term trend that show "moderate improvement". Based upon current 2012 to 2015 results current regulatory programs and actions are successful in reducing maximum ozone concentrations as 1-hour average data. This peer reviewer is in agreement with these findings.

4] TRPA has characterized compliance with the 8-hour ozone standard as "somewhat worse than target "with a trend that reflects "moderate improvement". The monitoring data for the period 1975 to 2014 supports the TRPA findings. A statistically significant downward trend is in place showing a 0.57% (0.0004 ppm)/year decrease in O_3 concentrations versus the 0.070 ppm California standard. This reviewer is in agreement with TRPA findings and conclusions.

Nitrogen Oxides-(NO_x)

1] This indicator currently relies on modeled emissions estimates of NOx from automobiles. Data are based solely on the California portion of the basin on what appears to be traffic count data collected in the Lake Tahoe Region (contributions from Nevada vehicles are not included). Results are reported in units of tons per day of NOx (average summer day) and are compared against a 5.6 ton/day threshold standard. These data have been trending downward since 1990 and are characterized in the report with "moderate improvement". Indicator status has been classified as "considerably better than target" based upon the following: 1) CARB (2015) estimated an average of 4 tons per day of NO_x or 71% of the 5.6 tons/day threshold. 2) 2014 annual average NO_x concentrations in ambient air 4.1 ppb or 14% of the strictest California standard. 3) 2014 highest 1 hour NO_x concentration of 27.9 ppb or 15.5% of strictest California standards. Confidence is classified as moderate. The latter applies to both indicator status as well as trends. This reviewer agrees with this assessment and the conclusions offered.

2] TRPA claims that existing federal, state and regional programs and actions are effective in controlling NO_x emissions and ultimately NO_x levels in ambient air. These claims on the effectiveness of existing programs and actions are based upon declining trends in CARB emissions estimates in combination with TRPA ambient monitoring data. This reviewer agrees with this assessment and the conclusions offered.

3] TRPA has made a recommendation to amend the threshold standard for NO_x and favor adoption of a numerical standard consistent with state and federal concentration standards. Measurement of NO_x concentrations in ambient air more accurately represents contributions from all sources and not just vehicle associated emissions as is the case with the current modeled NO_x values. Monitoring and modeling approaches

will continue for a period of 5 years (Note: this latter recommendation made by this peer reviewer in 2011). This peer reviewer endorses this proposed approach for monitoring the status and trends of nitrogen oxides in the future.

Visibility - Regional and Sub Regional Visibility Specific Comments

1] The visibility indicator program was revised based upon peer review comments offered for the 2011 TRPA report. The visibility threshold standard was comprised of nine (9) individual indicators in 2011. This has been reduced to four (4) indicators in 2015.

2] Regional visibility is characterized as "at or somewhat better than target" based upon the 3 year average data base 2012-2014. The regional visibility trend shows "little or no change". The long term trend over the period 1991 to 2014 suggests that programs and actions have been effective at maintaining and improving visibility. The only exception to this is negative influences on visibility in the Lake Tahoe Region directly attributable to wildfires outside the basin. Control of wildfires outside the basin is beyond the jurisdiction of TRPA. This reviewer agrees with these findings and conclusions.

3] The sub regional visibility monitoring program was restructured based upon the 2011 peer review and the DRI Chen 2011 report. A second sub regional monitoring site was established at Lake Tahoe Community College in 2014. Insufficient data currently exists to determine indicator status and trends. Sufficient data should be available from this new site in 2017.

PM_{10} Respirable Particulate

1] The annual average respirable particulate (PM_{10}) concentration indicator status was reported as "considerable better than target" with a moderately improving trend. Data have been collected since 1989 at numerous sites in the Tahoe Region. Data are compared to the annual National Ambient Air Quality Standard (NAAQS) of 50 ug/m^3 and a stricter California standard of 20 ug/m^3. The highest annual average PM_{10} concentration of 14.3 ug/m^3 was measured at the South Lake Tahoe site in 2014. This represents 71% of the stricter 20 ug/m^3 California standard. The trend line shows a decrease of 0.5 ug/m^3 or 2.5% per year relative to the same California standard. This improving long term trend suggests that programs and actions currently in place are effective at controlling PM_{10} concentrations. This reviewer agrees with this assessment and the conclusions offered.

2] TRPA offers a number of options for defining how the respirable particulate indicator will be evaluated in the future. The current method consists of use of the highest reading from the most recent monitoring period for comparison to the strictest California standard of 20 ug/m^3. This reviewer regards this approach as overly conservative since only the highest annual average concentration is used to compare to perhaps the strictest PM_{10} standard in the nation. Based upon review of the options offered by TRPA this reviewer recommends that the following be considered for evaluation of indicator status in the future:

- Average of all monitoring stations during the most recent monitoring period for comparison to 20 ug/m^3 standard.
- Number of exceedances during the current monitoring period.

3] The highest 24-hour average PM_{10} concentration indicator status was reported as "somewhat worse than target" with a trend characterized as "not changing". Data are compared to the stricter California standard of 50 ug/m^3. The highest 24 hour average of 50.8 ug/m^{3}was measured at the South Lake Tahoe site in 2014. This represents a value 102 percent of the California standard. The trend line shows an increase of 0.1 ug/m^3 pre year (or 0.2%). This trend characterized as reflecting "little or no change" suggests that existing programs and actions could have been more effectively implemented. This reviewer does not necessarily agree with these findings as they result from application of an overly conservative approach. The single highest 24 hour PM_{10}

concentration measured in the most recent monitoring year at any site is compared to a very strict California standard of 50 ug/m^3. The latter is 3 times less than the PM_{10} NAAQS of 150 ug/m^3. This approach needs to be evaluated and an alternative more realistic approach selected. TRPA, in fact, offers a series of options for evaluation of this indicator in the future. (pg. 3-45 Analytic Approach). This reviewer recommends that the following be considered for evaluation of indicator status in the future:

- Average of all monitoring stations during the most recent monitoring period for comparison to 50 ug/m^3 standard.
- Number of exceedance of 50 ug/m^3 standard during the current monitoring period.

$PM_{2.5}$ Specific Comments

1] 24 hour and annual average $PM_{2.5}$ concentrations are lower than threshold standard concentrations. The highest annual average $PM_{2.5}$ concentration for 2014 was 8 ug/m^3 measured at Tahoe City (67% of the stricter California standard of 12 ug/m^3). The highest 3 year average of the 98th percentile 24 hour $PM_{2.5}$ concentration for 2014 was 31 ug/m^3, also at Tahoe City (88% of the NAAQS of 35 ug/m^3). $PM_{2.5}$ data for the Bliss Site is much lower than Tahoe City. The highest average concentration at the Bliss Site was 40% of the 35 ug/m^3 target value during the 2012-2015 monitoring period. The trend based upon Bliss Site data indicates "little or no change". This based upon an increase in the 98th percentile 24 hour $PM_{2.5}$ concentration of 0.1 ug/m^3 (0.3% increase) annually. TRPA claims that programs and actions in place are effective at maintaining $PM_{2.5}$ concentrations below standards. The trend line, however, warrants continued attention and further corrective action if the trend continues to increase above the current annual value of 0.1 ug/m^3 (0.3 % increase). This reviewer is in agreement with these findings, conclusions and proposed course of action.

2] TRPA recommends that the specific definition for the $PM_{2.5}$ annual average indicator be reevaluated. Several potential options are offered. This reviewer agrees with this recommendation. Options recommended by this reviewer can be found in the PM_{10} section of this review (Comment 2).

Nitrate Deposition

1] A new nitrate deposition indicator has been introduced in the management standard category. The status of this newly implemented indicator could not be evaluated due to lack of relevant information. The TRPA has identified two (2) management standards to address nitrate deposition one under the air quality threshold standard and a second under the water quality threshold category. Available information regarding nutrient deposition suggested that there has been little or no statistical change in the amount of dissolved inorganic nitrogen deposited into the lake annually. These data suggest that the TRPA regional plan and stricter vehicle emission standards have not been effective.

2] Nitrogen deposition in Lake Tahoe is currently based upon water quality monitoring at a single location on the lake. TRPA recommends that the representativeness of this location be assessed. This reviewer agrees with this recommendation. In addition, multiple locations on the lake should be identified for nitrogen deposition monitoring on a routine basis. Water quality monitoring is the most effective means of evaluating nitrate deposition into the lake. Use of a single monitoring location is not representative of deposition rates lake wide.

3] Control of NO_x emissions sources within the TRPA air shed alone may not be effective in the further reductions of nitrogen deposition to the lake. Long range transport of nitrogen/nitrates from outside the Lake Tahoe air shed may also be a factor. The indicator status and downward trends for VMT (see Comment 4) supports this, for example.

4] Vehicle miles traveled, which is used by TRPA as a surrogate for both NO_x emissions and nitrogen deposition, have been trending downward since data capture began in 1981. TRPA has characterized this trend as representing "moderate improvement" as a consistent downward trend has been observed since 2009.

Indicator status is characterized as "at or somewhat better than target". (VMT estimated at 95% of 20,000,000 miles target). "The status and trends for VMT suggest that current programs and policies are mostly effective in reducing VMT". This reviewer agrees with this assessment and the conclusions offered.

Odor Specific Comments

1] TRPA's stated position in the TRPA 2011 Threshold Evaluation report was that policy statements that addressed diesel exhaust from vehicles be removed from the list of Threshold Standards. This reviewer at the time did not endorse the removal of these policies. In response to this the TRPA has retained policies, ordinances and programs and expanded the text in the 2015 Threshold Evaluation Report. This reviewer supports these actions.

2] This reviewer again recommends that a numerical Threshold Standard be put in place to allow for direct monitoring of the attainment status of the odor indicator, as well as, trends moving forward. TRPA, in fact, recommends that "applicable ambient air quality standards for NO_x, SO_2, CO and PM that are directly applicable to diesel engine emissions be used to measure attainment with the diesel odor standard". This reviewer endorses this proposed approach.

3] The odor indicator is restricted to vehicle emissions (e.g. diesel fueled) at present. Use of numerical ambient air quality standards for compounds associated with vehicle emissions such as NO_x, SO_2, CO and PM is appropriate. As ambient concentrations for these parameters trend downward odor incidents attributable to vehicle emissions should "track" accordingly. This reviewer endorses use of these standards as surrogates for odors attributable to vehicle emissions.

4] The TRPA report states that "it appears that state and federal measures and programs have been effective in reducing odor. In addition, state and federal actions in conjunction with adopted TRPA policies, appear to be sufficient in lieu of TRPA program support". (Effectiveness of Programs and Actions pg. 3-66). While these statements may be true there is no data in the odor indicator section to support these statements. As a result, the existence of odor programs, policies and ordinances alone do not serve as useful measures for monitoring odor indicator status and trends. Numerical standards such as those suggested in Comments 2 and 3 above should also be in place and applied in combination with policy statements for evaluating odor indicator status and trends in the future.

5] Odors attributable to sources other than motor vehicles and diesel emissions are not addressed in the odor section. The TRPA should consider a process to monitor and record odor complaints including use of an Odor Hotline. Odor events could be monitored and used to evaluate attainment with an actual numerical standard (eg goal equals a reduction in the number of complaints and/or violations from year to year). This could be accomplished through development of a complaint data base. Goals for attainment could be established moving forward consisting of net reductions in complaints and/or violations (actual number or %) from the prior reporting period. Complaint and violation trends could be monitored and used to evaluate the effectiveness of all existing odor regulations and policy statements. Enforcement responsibilities could be delegated to local law enforcement agencies such that additional TRPA resources would be limited. In a similar manner to the proposed approach for noise odor violations could result in warnings and fines levied against the offender.

Review by Dr. Jeffrey Marion

Virginia Tech University

Lake Tahoe Review for Chapter 1: Introduction

General Comments

I've completed my review of Chapter 1, Introduction. This chapter does an excellent job of putting the current land uses and impacts into a context of the substantially greater historic land use impacts associated with logging, grazing, and fire prevention. The role of research and science in evaluating and describing those impacts, and providing guidance to TRPA for more sustainable land-use practices is also well done. I do have some comments below about nomenclature related to the threshold standards being referred to as "targets." Hunters aim for targets but most environmental thresholds represent the "lowermost" condition boundary that managers seek to avoid crossing.

All the other material included are fine, I had no comments. It's all necessary and is concisely and clearly conveyed.

Specific Comments

Pg. 1-3, last paragraph: The Threshold standards are described as "targets" but aren't they actually defining the "lowermost" boundary of environmental condition acceptability. i.e., when the standard is exceeded a corrective action must be taken? If so they I would not refer to them as "targets;" managers are not trying to achieve the target but to "avoid" this lower threshold. The 1st paragraph on the next page also characterizes them this way ("achieving threshold standards"). Why wouldn't you say that managers are seeking to improve environmental conditions to avoid exceeding the threshold standards? Perhaps some of your standards are actually stated as targets that managers seek to achieve? If so then this complexity should be clarified so that readers are not confused.

Pg 1-5, item 2: Regression analysis – what I saw in the scenic quality chapter was simple regression (a correlation between two variables). However, multiple regression is a preferred and commonly used analytical method for modeling a complex array of factors to determine their relative influence on an environmental indicator of concern. However, I agree that this report should leave the causational analyses to separate research efforts, and simply summarize or report their most relevant findings, implications, and guidance. However, I suggest that this report, or a cited website, should contain citations of the full studies that are depended upon for providing relevant data. Links to the full reports and journal papers would also be beneficial and help to demonstrate the scientific basis for decision-making.

Pg 1-6, Numerical Standard: The water transparency standard is a good example of my comments above. That standard of 29.7 m is not a management target, avoiding it and achieving substantially greater water transparency is the object of management effort. You also don't want all waters throughout the lake to degrade to the standard – if some parts of the lake are substantially more transparent you want to keep them that way. Thus the report chapters should evaluate changes in resource condition over time (trends), not just "were the standards being exceeded."

Lake Tahoe Review for Chapter 2: Methodology

General Comments

I've completed my review of Chapter 2, Methodology. To begin, I'll note that the definitions included on page 2-1 of this chapter addressed my comments from Chapter 1 regarding nomenclature for standards as targets. The "Interim targets" appropriately refer to desired improved conditions that management seeks to "achieve." However, on page 2-2 under "Evaluation of Numerical Standards…" it reverts to calling the threshold standards "targets … of desired conditions" which most of them are not! I suggest that if your confusing and changing nomenclature confuses me then many of your readers will also be confused and may arrive at inappropriate conclusions…

There is some overlap between this and the preceding chapter.

Specific Comments

Pg. 2-4: OK, finally there is a description of three different types of numerical standards but the 1st two types are still not "targets" – they are or a "not to exceed" type. Even the range standards are not really a "target," only the Interim Targets are appropriately characterized in my personal opinion.

Pg. 2-5: I would replace the word target with the word standard in the color-based descriptions.

Pg. 2-7, Simple regression: A good concise explanation, though sample size problems and the possible need for transformations are not addressed.

Pg. 2-7, Estimating Interim Targets, last bullet: So if water transparency has a standard of 29.7 m and your current transparency in one portion of the lake is 100m why would you not make maintaining that high quality value your Interim Target?

Pg. 2-9, Confidence: This is fine, though the R^2 values could present problems. An OK "rule of thumb" but good judgement by respected scientists does not always agree w/the R^2 values you include.

Pg. 2-11: This is where you might consider linking to a website the lists the scientific citations and perhaps links to the best available and most relevant scientific reports and journal papers.

Lake Tahoe Review for Chapter 9: Scenic Resources

General Comments

I've completed my review of Chapter 9 on Scenic Resources. In general, I found this chapter to be a professional review of the scenic resources within the Lake Tahoe Basin. In contrast to the Recreation Chapter, the guiding statements for scenic resources do permit the development of effective evaluation indicators and standards and provide a sharp contrast to the deficiencies noted in the Recreation chapter. The indicators, monitoring methods, and analyses for this chapter are substantially better and provide an example of what the recreation section could look like with improved guiding statements. In particular, this chapter incorporated the kind of longitudinal (time-based) analyses, graphics, and statistical testing that are possible and necessary to inform management decision-making.

Some possible improvements include incorporation of discrete examples of ongoing scenic resource problems with photos to convey the most common challenges to readers. Similarly, a description of corrective actions with before and after photos would also be helpful. Without these readers don't gain a complete understanding of the challenges associated with the management of scenic resources. I note that the section on the Built Environment (pgs. 9-32 to 9-40) does include some of this type of material, which was very informative. Inclusion of photos showing new construction adjacent to the scenic evaluation critiques promotes dialogue and development of a shared norm or consensus on what the Lake Tahoe Basin community desires. These examples also clarify the management direction for building owners, architects, and the public.

Specific Comments

Pg. 9-2, Table 9.1: In contrast to Recreation, the Scenic Resources indicators and standards are indeed described to facilitate management decision-making and evaluation.

Pg. 9-3, last line of 2nd paragraph: The third word should be "route."

Table 9-2: Excellent table, particularly the division into the five Status & Trend classes. The Figures and map on the following page are also excellent – ***particularly*** the ratings by year in lower left showing the improving trend with statistical testing. This information is very helpful for interpreting the findings and a substantial improvement as compared to the Recreation chapter. I am a bit confused by comparison of this figure to the figure and data on page 9-7. One is % of units in attainment and the other is average travel rating – which is the "official" metric? Use of methods consistent with the U.S. Forest Service, which were developed by scientific research, is very helpful and lends credence to this section.

I was left wondering what the problems were for the units that did not attain an acceptable rating. Could that information be summarized and presented so that readers are more aware of the problems and possible solutions?

Table 9-2 and pg 9-11: Also excellent. Great to see the change in direction with improvements in shoreline travel route ratings beginning in 2002 following the Scenic Shoreland Ordinances (only possible by presenting the date by year). Excellent example of "adaptive management."

It would be helpful to know what some of the problems were, how they were resolved, and what some of the remaining problems are. Readers are left in the dark regarding "what's really going on." Some photos might also help characterize the conditions of concern or show the improvements through side-by-side before-and-after comparisons. Given the subjectivity of scenic quality assessments I'm surprised that some photos are not included here for illustration purposes.

Pg. 9-18 & 9-21: Same comments for the "Roadway Scenic Quality" and "Shoreline Scenic Quality" Ratings – could use some explicit examples and photos. I'm curious as to why they include both segment and point assessments and data. Might there be some comments to describe the advantages/disadvantages of both methods? Are both methods necessary or is one clearly better?

Pg. 9-23: It would be useful to describe some of the challenges or problems specifically, as well as some of the solutions that have been applied. Before and after photos would also be good to include as examples.

Lake Tahoe Review for Chapter 11: Recreation

General Comments

I've completed my review of Chapter 11 on Recreation. In general, I found this chapter to be a professional review of the recreational facilities and their development and management within the Lake Tahoe Basin. However, the guiding policy statements for the Tahoe Regional Planning Agency make their task rather challenging as described in my comments below. Until more clarity is established it will be difficult to make visitor and recreation management decisions and professionally manage increasing visitation. Reviewing the success of the various organizations that provide recreation facilities is equally difficult due to this guidance. I suggest refinements that specify the provision of optimal levels of appropriate recreational activities, high quality experiences, and the protection of Lake Tahoe's natural, cultural and historic resources. Further, I suggest incorporation of zoning so that some areas can be managed with substantial developments that can accommodate large numbers of visitor with limited resource impact, while other areas are managed with less development and visitation. Employing the ROS and VUM frameworks briefly described and included below can assist managers in the professional management of their lands and recreational infrastructure, activities, and

capacities (PAOT). Such frameworks are particularly beneficial when difficult and publicly salient management decisions must be made and justified.

Some improvements to this report could be undertaken currently of for the next report to include the element of time in the various tables and figures that would allow an improved understanding of long-term trends and management success. The report does include recommendations, which I support, to adopt improved standardized surveys of visitor satisfaction (e.g., the U.S. Forest Service Natl. Visitor Use Monitoring system surveys). Finally, I also suggest incorporation of monitoring to assess the resource conditions on and adjacent to the existing recreation facilities (recreation sites and trails).

Specific comments

The adopted threshold standards for recreation are statements of policy rather than numerical standards. 1) "…preserve and enhance high quality recreational experiences and provide additional access to the shore zone and other areas for dispersed recreational uses," and

2) "…establish and ensure a fair share of the total basin capacity for outdoor recreation is available to the general public." (TRPA 1982).

So TRPA recreation management is guided by these two statements? I'm not sure you are inviting my comment on those (they are likely beyond the scope of this report) but I do have some concerns. First there is no mention of natural, cultural, and historic resource protection from recreational visitation in either statement. Shouldn't there be? Second, statement 1 appears to call for perpetual recreational development of the shore zone, which is quite sensitive to visitor use impacts when compared to traditional "on-water" activities, or to recreational developments set back from shorelines. Statement 2 seems to address an equity issue but it's unclear to me who, other than the general public, is competing for a share of that capacity. More importantly, these statements are not written in a manner that allows a clear empirical evaluation of management success. The assessment asserts that "the Region remains in attainment with adopted recreation threshold standards," but the policy statements were not written to allow a clear evaluation of "attainment of standards."

The table on page 11-2 provides a good example and illustration. Surveys of recreation satisfaction in National Park Service units (where I generally work) are notoriously "favorable" – the great majority of visitors are highly satisfied. To be useful an evaluation would need to develop a sensitive and reliable assessment tool and apply the same tool through a qualified sampling process to yield comparable data that could be examined periodically to evaluate changes in visitor satisfaction over time. This table does not include survey data from earlier time periods so changes in satisfaction are unknown. I could not find any source information for these data in Appendix 1.

"The recreation element contains goals and policies that are intended "to achieve the intent of the thresholds over the life of the Plan by ensuring that recreational opportunities keep pace with public demand, that recreational facilities remain high on the development priority list, and that the quality of the outdoor recreational experience will be maintained." (TRPA 2012e)"

The statement above from the bottom of page 11-2, taken from the Regional Plan actually provides better and more measurable objectives. I still suggest inclusion of objectives related to the protection of natural, cultural, and historic resources. Now on page 11-3 I see reference to resource protection under Adopted Standards but it's not clear where this guidance is derived from.

Visitor surveys – I'm familiar with the U.S. Forest Service Natl. Visitor Use Monitoring system surveys. They are professionally developed, employ consistent sampling, and should provide good comparable information. It's unfortunate that their data was not yet available for inclusion. I recommend employing their data, extending them to non-USFS sites, and including statistical comparisons between consecutive surveys for future comparisons.

The data presented are from a range of sources and no information is included in Appendices allowing an evaluation of the survey methods, research design/methods, sample sizes and response rates, etc. in the future I suggest inclusion of such material in an Appendix. As noted, the best evaluation of attainment would include a comparison of findings from two or more identical surveys conducted several years apart – none of these include such comparisons. Having said that, visitor satisfaction does appear to be high for the south shore but is substantially lower for the north shore, and quite low for the S. Lake Tahoe and El Dorado County facilities (page 11-20 for the latter). Additional data on page 11-21 indicate that visitors desire more county parks, facilities, and trails and improves in the "ease of travel" for trails. Values in the 50's and 60's would not be suggestive of "in attainment" as indicated in the report.

While there is data documenting a trend in the acquisition of public lands it seems haphazard – a table revealing the various public land owners and their acreage by year with aggregate sums at the base would be helpful. Figure 11.1 does provide useful information for gauging the number of trail miles developed or improved trails by year – this is excellent though there should be a definition of what "improved" means or it should be omitted. Table 11.2 is also useful but data across time is needed. For example, how is the number of miles of roads with bicycle lanes and sidewalks expanding over time? Same for number of miles of street and mtn. bike trails and hiking trails by year. Improvements over time can't be evaluated without the incorporation of change over time. Table 11.3 does contain some time-based data by describing new projects but an evaluation of long-term changes is not possible.

Page 11-8, top: I see the recommendation I previously made regarding the use of the USFS surveys – good! If the timing of their survey is poor I suspect that some other entity could apply their same survey if you requested permission.

Pg 11-8, bottom: Great to see these 5 suggested strategies; they are excellent. In particular, the informational component is important. During my visit I did not realize that there was limited parking over at Emerald Bay and when we visited on a busy day we were unable to park there. I had examined the South Shore Bus service but discovered that it did not reach Emerald Bay. I did not discover the Emerald Bay Trolley service until just now when I did a Google search.

Pg 11-9: Would like to see a definition of "Improved" recreation facilities or have it omitted unless the improvements are truly substantial.

"The PAOT measure is an estimate of the number of individuals that a recreation facility or area can support at any given time. The PAOT allocations are used as both a target for desired recreation capacity, and a maximum limit to the recreational use that can be supported in an area."

Pg 11-11: The above statement seems like an entirely ineffective process. We use PAOT determinations to identify the maximum number of visitors at various destination spots within National Park Service units. Almost always the number of visitors at those maximum capacities allow a degree of crowding and "less than desirable" experiential conditions. They should not be used as "a target for desired recreation capacity" in the sense that you are actively managing to increase use to that "target." Did I misread this?

I've tried to understand how the PAOT system is being applied but the report is confusing on this topic. For example, it discusses the Eagle Falls trailhead and its 36 parking spaces. However, if a bus drops visitors here the design capacity of the parking lot becomes irrelevant when determining PAOT. I suggest application of a Recreational Opportunity Spectrum (ROS) system for planning and management purposes, and of a Visitor Use Management (VUM) framework for making decisions regarding the management of visitor use and carrying capacity. This are briefly described below. The ROS system allows land managers to develop and manage different zones to achieve differing resource protection and recreational objectives.

Recreation Opportunity Spectrum (ROS) is a system for classifying and managing recreation opportunities based on the following criteria: physical setting, social setting, and managerial setting. The combination of the three criteria results in six different ROS classes: Primitive, semi-primitive non-motorized, semi- primitive motorized, roaded natural, rural, modern urban. It recognizes that a common activity like hiking can provide very different experiential outcomes depending on where that activity occurs, e.g., in a suburban setting vs. a wilderness setting.

Websites:

http://www.fs.fed.us/cdt/carrying_capacity/rosfieldguide/ros_primer_and_field_guide.htm

http://www.fs.fed.us/cdt/carrying_capacity/gtr098.pdf

Visitor Use Management (VUM) is the proactive and adaptive process for managing characteristics of visitor use and the natural and managerial setting using a variety of strategies and tools to achieve and maintain desired resource conditions and visitor experiences. Developed and just released this week by a federal Interagency Council, it provides improvements over prior frameworks used by the U.S. Forest Service (Limits of Acceptable Change, LAC) and the National Park Service (Visitor Experience & Resource Protection VERP) that it will likely replace.

Websites: http://visitorusemanagement.nps.gov/Home/About, http://visitorusemanagement.nps.gov/Content/documents/Position_Paper_Volume_One.pdf

Pg 11-13, under Monitoring Approach: I concur with the report statements about the need to improve the PAOT system to account for a wider range of criteria. I'm a scientist who specializes in recreation ecology, a small field of study that examines the environmental impacts of recreational activities, and works with managers to avoid or minimize these effects. Our studies routinely find that the same amount and type of visitor use can have widely different impacts depending on an array of different factors. For example, hiking traffic on a paved trail is basically inconsequential (no measurable impact) but that same number of visitors hiking along an undeveloped shoreline lacking a trail could have significant impact.

I don't see any condition assessment monitoring of recreation sites or trails. For example, is this infrastructure sustainably designed and managed and what condition is it in? Poorly designed or overused trails and recreation sites could have soil loss or visitor activities and resource impacts extending outside the designed "footprint" of intended use. I suggest adding future assessments and monitoring to include such monitoring and I can be of assistance in recommending appropriate methodologies.

Pg 11-13, under Modification of the Threshold Standard or Indicator: I also concur with the report statements here about the need to clarify guiding objective terms (as previously noted). I still do not understand who other than the "public" is competing for the Lake Tahoe Basin resources. Does it mean the public traveling with a commercial party (on a ferry, renting a commercial boat) vs. unaffiliated visitors or private land owners? This needs to be clarified.

Review by Dr. Robert Naiman

University of Washington

Summarized below are evaluations and comments on the information, analyses, results and recommendations contained in the 2015 TRPA Threshold Evaluation Report for the general Introduction, the general Methodology, and the Fisheries and the Vegetation chapters. As requested, the documents were evaluated relative to my expertise of the subject matter, publicly available information, and commonly accepted scientific methods and practices. I provide critical evaluations that identify deficiencies and errors, as well as point out positive aspects. Where appropriate I have made suggestions for objectively addressing most of the identified deficiencies or errors.

General Introduction - *Does the Introduction Section provide sufficient background information necessary to understand the purpose and scope of the Threshold Evaluation Report?*

The Introduction contains considerable information but does not present a compelling rationale for the importance of the Tahoe Regional Planning Agency (TRPA) program. The first page needs a concise paragraph giving the purpose, objectives and rationale for this report. In other words, provide a compelling paragraph to communicate to the readers why this report – and the program – is vitally important. As well, on Page 5, inform readers about the mechanisms or processes driving the observed status and trends. The activities are not convincing if the document is limited because analyses of observed status and trends – and causes and effects – are often not possible because the "*Investigation and presentation of the mechanisms and processes driving observed status and trends is beyond the charge given to the four-year threshold evaluation*". This may be true but it is not what readers want to hear. The Introduction could be improved by crafting this difficult truth in a more positive light. How is the program going to identify the long term status and trends, and more importantly, what does the program plan to do about them, if anything? Likewise (also P. 5), if regression is not adequate, why not use the appropriate statistical/experimental procedures? Statement like this one leave readers wondering about the reliability of information presented.

Another general issue with the Introduction is that key words related to standards and confidences are not quantitatively defined. For instance, vague words such as *high quality, natural, moderate, and low* all have different connotations, depending on how they are used and perceived by individual readers. While this is a common problem in general reports it has been adequately addressed by several groups, notably the Millennium Ecosystem Assessment (http://www.millenniumassessment.org/en/index.html). The authors may want to review the approach they adopted.

<u>Specific Comments:</u>

P. 9: Under Monitoring and Analysis provide general information on data storage, determination of quality, and the general availability of data.

P. 9: Inform readers about how an "*indicator*" can be "*out of attainment*". Otherwise, this is a confusing statement.

P. 10: The recommendations seem to address mostly scientific approaches. If the recommendations are to fully engage readers and the public there need to be recommendations that also address land use or other human-driven activities. Are there general recommendations of this type that can be added to the Introduction?

P. 10: Here is an opportunity to improve the overall program. What will it take to have management standards that are quantifiable? This seems essential for a fully functioning program.

There are a few additional comments and suggestions on a marked copy of the Introduction, which is attached.

Are statements of fact presented in the Introduction Section sufficiently supported with appropriate references or original data and analysis?

In general, yes. The Introduction provides references to published articles and reports as well as websites. Nevertheless, a link to a comprehensive database of reports, articles, websites and other resources would be highly beneficial.

General Methodology - *Are prescribed approaches for determining the status and trend of indicators relative to adopted standards clearly presented and appropriate?*

While the program is admittedly complex and evolving, this chapter raised many fundamental questions in my mind as to the veracity of the overall approach. The general Methodology perhaps requires a more in-depth discussion (and resolution) among the project leadership as to what the program is truly trying to attain, and whether or not the approaches employed will allow the program to reach the ultimate objectives. For instance (P. 1), it will be important for readers to understand what scientific or policy mechanisms are used to determine the Threshold Standards, and if they are ecologically or socially meaningful and defensible.

The chapter, in several places, suffers from poor sentence construction. I recommend that it be carefully edited for clarity. Specific comments (and some questions) can be found on the attached document.

Is the prescribed approach for determining the level of confidence in status and trend determinations clearly presented and appropriate?

While environmental characteristics vary over space, time and land use, using the current year's value can be misleading depending on that variation. I suggest reporting SD's and ranges, where appropriate. Further, are the evaluations of trends statistically significant (P. 6)? If not, then how is a trend determined?

On P. 7 there is a detailed explanation of regression analysis. More importantly, it appears that PCA's and other more advanced statistical analyses are not being used. Why? The approach being used here (regression analysis) is a good start but may not be the best approach for determining complex trends and causalities. A defensible explanation is needed as to why other statistical approaches are not used.

P. 8: What is the scientific rationale for using 2.5% and other percentages?

Using the intersection of the horizontal line with the y-axis to estimate an interim target for a specific indicator may not be always a good idea from an ecological perspective (P. 8). For instance, consider water quality (e.g., N, P) and its influence on fish carrying capacity. While it may be a good idea to reduce N & P in culturally eutrophic situations, it may not be a good idea in other situations to reduce them to zero (or very low levels). Nutrients are required to sustain healthy populations. Perhaps something needs to be said here about attainment of standards. Additionally, how is a quantifiable target set in a situation where the objective is to slow the rate of change away from attainment? Readers will want to understand this process.

The section on confidence in trend determination (P. 9 and 10) raises a number of questions. First, it is important to define "*Trustworthy Data*". Second, what is the scientific rationale for using the values and cut-off points (e.g., >0.75, and others), especially when data are sparse? It seems misleading without strong ecological and social rationales.

For the overall confidence determinations (P. 10, Table 2-4), I recommend that it would be prudent to error on the side of environmental caution and use the lowest rating, especially when one is high and the other low (or could not be determined). Is this philosophy adopted by the program?

The Management Standards and Policy Statements (P. 11) do not include assessments of trends or confidence, which is understandable. However, do they include specific recommendations on what needs to be done to attain the management standards and policy statements? The statements should at least provide explicit examples of what needs to be done, especially if readers are to have confidence that the program has thoughtful and committed leadership.

Are the approaches prescribed to determine an interim target and attainment date for an indicator reasonable, given data and funding limitations?

See comments above. No specific information was provided on data and funding limitations in this chapter; therefore, it is not possible to answer the question.

In evaluating the chapters on Fisheries and Vegetation, I used guidelines provided to reviewers by the Program coordinators:

Introduction - Does the Introduction Section provide sufficient background information necessary to understand the purpose and scope of the Threshold Evaluation Report? Are statements of fact presented in the Introduction Section sufficiently supported with appropriate references or original data and analysis?

Methodology - Are prescribed approaches for determining the status and trend of indicators relative to adopted standards clearly presented and appropriate? Is the prescribed approach for determining the level of confidence in status and trend determinations clearly presented and appropriate? Are the approaches prescribed to determine an interim target and attainment date for an indicator reasonable, given data and funding limitations? Do the authors clearly articulate the sources of uncertainty in the analysis?

Indicator and Indicator Reporting Category Evaluations – Is the write-up associated with each indicator evaluation clear and complete? Are the analytical methods appropriately applied in the determination of an indicator's status, trend, and confidence? Are there other or different analyses you would recommend to evaluate indicator status or trend? Are statements of fact or conclusions supported by appropriate references or original data and analysis?

Recommendations - Are recommendations appropriately supported by report findings and conclusions? Are proposed recommendations related to policy and on-the-ground project implementation supported by best available science? Are remaining uncertainties clearly articulated?

Fisheries

The title of this chapter is misleading. There is little factual information on the status of fish populations and none on the commercial/recreational fisheries. Instead, the chapter is focused on a set of the more important aquatic habitats potentially available to fish. The TRPA should note that Fisheries (the plural form of "Fishery") is defined as 1. a place where fish are bred; such as a fish hatchery. 2. A place where fish or shellfish are caught. 3. The occupation or industry of catching, processing, or selling fish or shellfish. 4. *Law:* the right to fish in certain waters or at certain times. This chapter should be re-titled to reflect actual contents.

While the chapter contains useful information that, for the most part is adequately acquired, the "*Fisheries*" program was a disappointment overall. This assessment relates to the emphasis on coarse habitat changes rather than fish population characteristics, the lack of clarity (quantitative definitions) in evaluation categories, the limited or non-existent (e.g., in-stream flow) monitoring, and the limited amount of restoration attempted. The chapter raised many basic questions and concerns, and I've listed the more important ones here and in the attached mark-up of the chapter.

P.1: It should be made clear to the readers as to why the fish and fisheries are ecologically and socially important. What do they tell one about the state of the Tahoe ecosystem? Do they have recreational or commercial significance?

P.1: The second paragraph could be moved to a section on "*History*". Its present position detracts from the overall flow of the text; the rationale and goals of the chapter should occupy the lead paragraphs.

P.4: How are *excellent, good, and marginal conditions* defined? Ideally, it should be by the responses of fish population characteristics over time.

P.5: Lake Habitat is, in reality, littoral habitat of undetermined depth (30m, 50m?). Are the pelagic or deeper (profundal) habitats considered in the status and trends evaluations? If not, why not? Are there valid ecological reasons for not considering them?

P.5: How was the management target of 5,958 acres determined? [Note: 5,948 acres are used elsewhere in the report] Are there ecological or social rationales for fixing a very precise acreage? How many acres are potentially available for restoration or management? The indicator for lake habitat does not seem to have established a depth. Is there a "*cutoff*" depth for the acreage?

P.7: Provide an ecological rationale for establishing the "*prime*" size categories for substrates.

P.8: Are invasive plants an issue in the littoral zone? If so, are they detected by remote sensing?

P.8: If aerial mapping techniques and approaches continue to change, how will one determine if there are valid trends in littoral habitat? This issue needs to be given high priority by researchers.

P.9: Since "*there is insufficient data available to assess the effectiveness of individual programs or actions in maintaining lake fish habitat*" shouldn't the program implement research or monitoring to directly link them? Otherwise how does one know that the programs and actions - operated at great effort and expense - are effective with respect to habitat carrying capacity and fish population responses?

P.9: The recommendations for lake littoral habitat monitoring are reasonable and should be given serious consideration for adoption.

P.11: Scientific defensible, quantitative definitions of *excellent, good and marginal streams* are needed.

P.14: Only eight reference sites have been established. I doubt that this is analytically sufficient. Are the reference sites established on various physical types of streams considered to be in "*excellent*" condition in order to measure departures from natural conditions? Having a scientifically defensible comprehensive strategy is essential if changes (or no changes) are to be detected in a reasonable period of time.

P.15: Provide ecological rationales for using the numerical ratings and their thresholds for streams in the TRPA program. Otherwise they have no meaning.

P.16: Evidence needs to be provided to demonstrate that physical habitat characteristics (which are not identified) and "*break points*" established for streams elsewhere in California are applicable to streams in the Tahoe Basin. As pointed out later in this section of the report, additional physical characteristics and climate play important roles in stream "*health*". Were they included in establishing the category designations for the Tahoe streams?

P.16: The statement about categorization being mislabeled based on "*traditional*" assessments requires considerably more explanation. What does this mean in terms of scientific conclusions as well as management and policy decisions?

P.17: It is hard to accept the "*high*" confidence in the status of streams (and their fish populations) in the Basin. A very large percentage are rated as excellent - implying the fish populations are in good condition - but there

are many culverts blocking movement, few native fishes, and the population dynamics are poorly monitored. Biotic integrity, as measured by the techniques used, may be high but there is little evidence that fish populations are in good or excellent condition.

P.18: The statement that streams and fish habitat have been substantially restored or enhanced through the Environmental Improvement Program, is hardly true with only 5 miles treated. This is a gross over-statement and should be modified to reflect improvements that can be documented. Later it is admitted that "*the positive effects of these projects on overall stream health have not yet shown up.*" The reasons provided for not seeing positive effects are valid.

Additionally, it is disappointing to see that fish passage improvement over 5 years resulted in only an additional 7.66 miles of available habitat (1.53 miles/yr). This is a very low rate of improvement and attempts should be made to greatly accelerate restoration in future years, especially since the blockage points are known.

P.19: A recommendation is needed that addresses the slow rates of habitat restoration and removal of passage barriers.

P.19: The recommendation that the threshold standard or indicator should clarify its intent to reflect the overall biotic integrity of the stream, rather than just fish habitat alone, is a good one – and should be seriously considered by the TRPA.

P.21: How are in-stream flow targets related to fish and other aquatic/riparian characteristics and ecological needs? This is a vitally important relationship and needs to be addressed directly via research and experimentation.

P.21: Without monitoring it is impossible to determine if in-stream flows are being provided. It, therefore, is premature to state that "*the threshold standards are determined to be implemented and in attainment*". The lack of monitoring for compliance is a major oversight.

P.22: How do groundwater (GW) extractions affect stream flows? I would suspect that large GW extractions in a limited basin would have significant impacts on in-stream flows, especially during periods of drought.

P.22: While it is laudable to have a statistical model for trout in-stream flow needs, other species have different flow requirements. The effort would be more robust if contrasting ecological water needs were considered (e.g., riparian plants, amphibians). As well, winter flow and ice conditions are extremely important in influencing biotic components. No information was provided on winter conditions.

Further, the Tracy and Rost investigation is 13 yrs old (and a non-peered reviewed internal report). How have conditions changed since then? Would the recommendations be the same in 2016? As a non-peered reviewed report, considerable emphasis is being placed on its veracity; it may be time to re-evaluate the issue.

P.23: The suggestion that "*... the need to establish minimum flow standards for individual streams may not be feasible or warranted and should be re-considered*." is not warranted, especially with the apparently widespread pumping of GW.

P.23: The recommendation for monitoring needs to be modified, especially since there is no monitoring to determine if in-stream flow criteria are being attained. Further, it has not been demonstrated that the existing flow regimes are fully beneficial to native fishes or other important community components (e.g., riparian plants). This is a serious programmatic oversight that needs to be rectified.

P.25: It is hard to reconcile the determination that so many of the streams are being rated as "*excellent*" habitat when restoration of LCT populations appears to be limited by stream habitat, blockages, and non-native competition. The authors should reconsider this conclusion.

P.26: It has been 4 yrs since LCT were introduced into Lake Tahoe. Has it been successful? What are the results in terms of recreational fisheries?

P.26: An additional recommendation might be to expand the number of sites being restored for LCT.

Vegetation

There is a substantial amount of information provided in this chapter spanning a broad range of spatial scales (from individual communities to the basin scale and, at times, beyond the Tahoe catchment). As such, the chapter has the potential to eventually become a "go to" source for future investigations. However, in order to perform that function or service, it will need to identify the data sources and ownership, evaluate the validity of data sets, and provide instructions on how data sets can be accessed.

The chapter could benefit from having a section (or summary) on lessons learned as well as a section on plant communities that have been overlooked or ignored for various reasons. In the case of the latter, it is surprising that phytoplankton are not considered, especially with the emphasis on lake clarity, and that non-native species and understory species seem to have been relegated to the back shelf. As well, the understory vegetation is vitally important in terms of biodiversity and ecological processes. Has it been quantified for species, non-natives, and other important ecological characteristics in the riparian and upland forests? These plant categories help underpin the long-term ecological vitality of the Tahoe basin and should be receiving appropriate attention.

It was pleasing to see that so many of the recommendations addressed the need for establishing quantifiable objectives, including timelines for attainment of objectives. For many plant communities, it will take decades to achieve the objectives but it is vitally important to have quantifiable targets along the way, as well as an Adaptive Management (AM) framework to guide management actions. My sense, from reading the chapter, is that the program has not yet attained a working integrated system that would allow quantifiable monitoring to inform a viable AM framework, but progress is being made. Perhaps it is covered elsewhere in the report but a section on the AM framework would have been greatly appreciated. I hope that programmatic progress is fast enough to keep up with the ever-emerging environmental issues and land changes.

In addition, I have quite a few concerns and comments about the chapter contents. While I found the individual sub-sections to be organized and at times informative, I also found many of them to be frustratingly limited in their temporal coverages and levels of monitoring. Here are some important comments, reactions and suggestions:

The current Introduction is largely "Background" information that could be moved to a new section. An Introduction should address the environmental issues - in this case, those associated with the vegetation - and provide clear objectives for the vegetative section of the TRPA report.

In the Table (P. 6-5 and 6-6, as well as elsewhere in the chapter), how were the percentages for relative abundance and other categories determined to be ecologically important and viable for the long term? As well, is there an Adaptive Management (AM) process in place to provide timely adjustments to specific program objectives? If not, there needs to be one.

I'm very surprised that suitable aerial photos do not exist before 2009 (see P. 6-13), at least for portions of the Basin. These would be highly valuable in determining trends, at least for some vegetative categories. What are the reasons why earlier photos cannot be used? In my experience, this is a very unusual development (see comment below on riparian vegetation).

The discussion of disturbed vs. undisturbed wetland/meadow vegetation in proportion to the total undisturbed vegetation for the Basin (P. 6-18) is very informative. It provides a strong overview that readers will appreciate.

P. 6-19: I agree that the indicator does not measure the relative condition of meadows and wetlands or their ability to support various ecosystem services or attributes. As stated, the standard should be assessed against best practice for the establishment of standards and indicators for monitoring and evaluation, and amended as necessary to improve the evaluability of the standard and the information it provides for management.

P. 6-20: If data for riparian vegetation are not reliable then they should not be shown. I'm surprised that calibrations between sets of techniques, photos, and so forth are not available (or were not done). Usually, as a matter of course, the calibrations are done and are well documented.

Riparian Vegetation: It is not clear how the TRPA arrived at a 4% target as being ecologically important for the riparian vegetation cover. It seems that there are better criteria that could be derived from aerial photos, such as the continuous distribution along streams, riparian widths, tree age distribution (or height), and other parameters that give an indication as to whether the riparian zone is structurally sound.

P. 6-22: Quantifying changes in condition are paramount for an effective monitoring program, and I agree with this recommendation for riparian vegetation (as well as for all vegetative categories).

Recommendations for the Shrub Vegetation Type: I do not agree with the recommendation for the Analytic Approach (P. 6-26). Having multiple estimates over relatively short (5 year) periods eventually provides some confidence that data are realistic (or not). It is also easier to establish trends, if they exist, with multiple points over several decades. Otherwise, changes can be interpreted as due to chance or other perceived drivers. Nevertheless, the recommendation for Monitoring is excellent, as well as the recommendation for modifying the Threshold Standard or Indicator. I support their adoption by the TRPA.

There are a number of concerns about the section on yellow pine and red fir forest (P 6-28):

1. Why the change in diameter reporting?

2. It is hard to keep believing this argument (changing mapping techniques/resolution) for why data are not reliable. A good remote sensing team would have made sure that, despite different techniques being used, the vegetation data would have been comparable.

3. Map is too small to read.

4. Once again, what is the ecological rationale for setting the percentages (P 6-29) for Adopted Standards?

5. Is active planting being considered to meet the Target Attainment Date for yellow pine and red fir (P 6-30)?

6. Why is "*The less than 10.9-inch dbh definition of small trees … <u>thought</u> to better represent the intent of the threshold standard (TRPA 2012b*)" (p 6-30)? As stated, it seems highly arbitrary.

7. Under recommendations, I do not agree with the proposed change in Analytic Approach; see previous comment on this topic. On the other hand, the revised monitoring approach and the modification of the threshold standard for yellow pine and red fir are good suggestions and should be adopted (P 6-31).

I was pleased to see a section in the vegetative report on the juxtaposition of vegetative communities and age classes (P 6-32). This speaks to a well-coordinated effort, besides being a highly functional activity as well. Nevertheless, a couple comments:

1. The map does not illustrate trends. Are there trends? Also, it shows fuel reduction treatments; not juxtaposition of vegetative communities or age classes.

2. The photos (P 6-32) do not illustrate vegetative juxtaposition. As well, they are repeated in the section on fuel reduction.

Bailey Land Capability System: The report requires a list (or a link to a website) giving the dominant vegetation or water types for each land capability classification (P 6-36). Further, how does the program address the proliferation of non-native species? How ecologically important are they in the Tahoe region?

Stream Environment Zones (SEZ): It appears that the SEZ program is structured around relatively recent environmental conditions (P 6-39). I suspect that recent conditions and perhaps spatial extent (especially near the lake) are significantly different from those before the arrival of Europeans. Early land activities most likely resulted in substantial sediment routing and deposition near the lake, local channel downcutting, removal of natural wood jams, and other environmental changes of long duration. Will the program attempt to restore the overall environmental conditions similar to those in a natural setting?

1. Why are there no monitoring partners and no monitoring as there is for upland vegetation? It seems that these would be essential.

2. The removal of conifers from SEZs seems counter-productive as many riparian areas have conifers. They eventually supply large wood to streams/wetlands that provides habitat for centuries.

3. How are the "improved" conditions determined SEZs determined if there is no long term monitoring (P 6-41)? This does not make sense. However, later on the same page the author's note that monitoring and modification recommendations are needed; these should be adopted.

Appropriate Management Practices (P 6-42): Identify the specific management practices. Otherwise one cannot judge if they are appropriate (or relevant) for the management objective.

1. Where would one find the evidence supporting this statement/determination on the status of implementing appropriate management practices (P 6-43)? Who made the determination? This topic requires more substance than is currently presented.
2. The recommendation for modifying the Indicator for Appropriate Management Practices is needed (P 6-44); it is one that should be adopted. Is there a way that this could become a quantitative objective with measurable outcomes?

Uncommon Plant Communities (P 6-45): Many of these appear to be places rather than plant communities. Mixing places and communities causes problems in conservation/restoration and, above all, retards learning from conservation/restoration actions and the development of conservation/restoration principles.

Deep-water plants (P 6-49): The authors provide a much needed recommendation for monitoring.

For the Upper Truckee Marsh (P 6-51) it is important to say why the condition is degraded; many readers will not recognize the degraded conditions from the photos because they have become such a common sight – many people see these as natural features.

Upper Truckee Marsh: Here and elsewhere in the program, a qualitative assessment of a plant community has little value for evaluating the ecological condition (P 6-52). I strongly suggest that this indicator become quantitative.

Upper Truckee Marsh: Should there be a recommendation that addresses non-native species (P 6-56)?

The report should provide an Introduction and rationale as to why the individual marshes are addressed separately. Could the earlier section on marshes and wetlands act as an Introduction (overview) to these individual sections by giving the overall trends? As presented, these seem to be appendices to the main report. To follow on, after reading several of the following sections on individual sites it is clear that there is a lot of redundancy by not combining the marshes/wetlands into one category. I suggest that the salient points be summarized in a Table, if only to make the document more readily comprehensible.

P. 6-67: Part of the conservation significance score was confusing. If burrowing rodents and beaver are native species, then why they should detract from the conservation significance score. It seems as though the scoring is being done by researchers biased toward plants ...

Freel Peak Community: The recommendation to attain or maintain a threshold is a good one (P 6-83); it seems that it may be the best action since the drivers of change in this community are global.

Rorippa (P 6-85): Provide the correlation coefficient of the relation between sites occupied and lake level in the Figure caption. It may be informative, and very useful.

P. 6-91: Where is the AM framework described for TYC?

Draba (P 6-92): This is a well written section. The authors make a strong case for the need to protect populations of Tahoe draba. Similar ecological rationales should be crafted for other species and habitat types.

Galena Creek Rockcress (P 6-103): Remove lines in the figures; with so few data they are misleading, especially without statistical significance.

Galena Creek Rockcress (P 6-105): The authors provide an important monitoring recommendation. As well, if there are only 2 verified populations, I agree that the target should be modified (P 6-106).

The section on Late Seral and Old Growth Forest Ecosystems (P 6-107) is important and probably should appear earlier in the chapter. As well, there should be a similar section on riparian forests. Why are riparian forests not receiving equal consideration? While they are not as spatially extensive they perform vital ecological functions well out of proportion to their land cover. Other concerns with this sub-section:

1. (P 6-109): Explain how 55% was determined to be the target and, as well, why that percentage is ecologically meaningful. Also, the same needs to be explained for the zonal distribution of percentages.
2. (P 6-110): The analytic approach does not seem to consider mortality as the stands undergo natural thinning. How are mortality rates, by species, incorporated into projections? As well, rather than use a single growth rate, it would be prudent to use a conservative range of observed growth rates to show the natural variability in transitions to "old growth" stands.
3. (P 6-112): The recommendation on monitoring does not address changes needed in monitoring. In fact, the paragraph does not address monitoring.
4. (P 6-112): The authors provide a logical recommendation for the subalpine zone that should be seriously considered.

Editorial Suggestions on the Vegetation Chapter

The key (Fig. 6.1) should appear before the symbols are used in the report (i.e., before the previous Table).

Define or spell out SEZ (P 6-22).

P. 6-46: Not sure what this means: "... two cycles of monitoring data." Also, on the same page, define "deep-water".

Deep-water Plants (P 6-47): The contour lines cannot be seen on the map.

Do you mean Grass Lake rather than Hell Hole (P 6-77)? See also last sentence on this page.

P. 6-99: Define SWE and SWE = 0.

Old Growth Ecosystems (P 6-108): The color key for the figure is missing two categories.

Review by Dr. Barry Noon

Colorado State University

I have thoroughly read the required chapters to enable me to provide an informed review of the wildlife monitoring program. The Lake Tahoe Regional Planning Agency (TRPA) is to be commended for providing some assessment of the status of wildlife with the Basin and to have established thresholds—that is, target values—for the monitoring state variables. However, my overall evaluation of the wildlife monitoring program is not favorable. Very few of the essential ecological and design components required of an environmental monitoring program are apparent in the wildlife program. I fully recognize that these limitations may be a consequence of severe funding constraints.

In order to make my comments as useful as possible to the TRPA, I have structured my review to address: 1) what I perceive to be a deficiency in the existing program; and 2) what I see as a possible remedy. Most of my substantive comments focus on Chapter 8 Wildlife. My comments on Chapters 1 (Introduction) are less detailed because this chapter is mostly background material to provide context and a general description of methods and approaches. Substantive comments on Chapter 2 (Methodology) are mostly contained in my review comments on Chapter 8 (Wildlife) because this allows me to be more specific.

Many of my general comments on Chapters 1 and 2 are repeated in my comments on Chapter 8 in the context of the specifics of wildlife monitoring and assessment.

Chapter 1 (Introduction):

General comments:

The identification of threshold standards in this chapter is very good. Establishing quantitative threshold values for measurable indicator variables is an essential component of any monitoring program. These critical values serve as decision points and link monitoring outcomes to the management decision-making process.

To make the discussion in this chapter more consistent with the published literature on monitoring ecological systems, I suggest that **ecological thresholds** be more clearly defined and distinguished from **decision thresholds**. Threshold responses in ecological systems represent rapid, non-linear changes in one or more key state variables—variables that characterized a key condition or output of an ecological system—in response to environmental stressors. Often thresholds can only be identified after they are surpassed and such threshold responses may signal effectively irreversible changes to the system. Because of this, it is important to establish decision thresholds that are anticipatory of an approach to an ecological threshold. This allows the possibility of a management response (e.g., reduce the magnitude and extent of the stressor) prior to passing an irreversible ecological threshold.

Specific comments:

- Establishing numerical standards for indicator variables is very difficult. That is, what value of a monitoring state variable implies an approach to an ecological threshold? To answer this question requires prior information on the natural variability and rate of change of the state variable and what values (rates) are unacceptable.
- The use of the reporting icon is very useful and easily understood by the public. The components of the reporting icon—status, trend and certainty—are logical but obviously require more detailed description when they are applied to specific resources (e.g., air quality, wildlife, etc.).

- Monitoring and analysis "approaches" are essential components that require a detailed explanation. This chapter is not the appropriate place for this discussion. However, in my review of Chapter 8, I found the discussion of sampling design components to be almost totally lacking or poorly described and justified.

Chapter 2 (Methodology)

General comments:

Most of the essential components of an ecological status and trend monitoring program are discussed in this chapter. However, the devil is in the details. For example, selecting indicators (monitoring state variables) is very challenging because managers are attempting to make inference to the integrity of a complex system (Lake Tahoe Basin) in terms of a small number of measured attributes. This requires careful consideration of what an indicator is actually measuring and, most important, if the value of that indicator is providing information beyond its own measurement.

The section on evaluation of indicator trend is very important. I only reviewed this topic in the context of Chapter 8 (Wildlife) and do not know how generally my comments apply to other resources that are monitored. The key point is that the form of the statistical model that relates the response variable (value of the indicator at time t or location i) to the predictor variable(s) (e.g., time, various environmental variables) must be carefully considered. In the case of wildlife population abundance, for example, default to simple or multiple regression is often not the appropriate statistical model. Modeling dynamic changes in the values of an indicator over time requires a clear understanding of those dynamical changes—for example, does the indicator change in an additive or multiplicative fashion?

In my opinion, the discussion in this chapter on confidence in status and trend estimates needs to be more strongly grounded in statistical theory. The foundation for reliable inference is in clearly defining the **sample frame** (target population) and how this frame will be sampled in order to allow reliable inference from the sample to the frame (i.e., the sample design components). Design components include the size of the sample units, the number of sample units, how units will be probabilistically selected, and the resample interval. Importantly, decisions on the number of sample units cannot be made without first identifying the magnitude of change in the indicator variable that is ecologically meaningful. Further, discussion of "confidence" should be framed in terms of **accuracy** of the estimated trend (**precision** + **bias**^{2}). One way to do this, in terms of the simple linear regression model, is to compute a prediction interval to the fitted regression line.

Chapter 8 (Wildlife)

Fundamental Deficiencies of the TRPA Wildlife Monitoring Program

The goals of the monitoring program and the target area to which these goals apply must be clearly stated. I interpreted the first few paragraphs of Chapter 8 Wildlife to have an implicit goal of sustaining all wildlife (vertebrates?) within the Lake Tahoe Basin. I believe this to be an appropriate **target population** (the vertebrate species assemblage) and **sample frame** (Lake Tahoe Basin).

Monitoring any complex system, such as an assemblage of wildlife species that may exceed 300 vertebrate species (i.e., Lake Tahoe Basin), requires a surrogate-based approach. That is, all species of interest cannot be directly measured. This is driven in large part by the pragmatic reality of limited budgets. However, if the goal is to make inference to the status and trend of the full wildlife assemblage within the Basin (which I believe should be the goal of the TRPA), then the small number of species selected for monitoring should possess a number of key properties including:

- Knowledge of their status and trend should tell you something beyond their own measurement
- They should have sampling properties such as adequate abundance and broad spatial distribution in order to estimate change in their status and trend with sufficient statistical power
- In combination with the other species in the surrogate set, the species should be **complementary** in information content with species already in the surrogate set and add to the **comprehensive** nature of the surrogate set (optimally, the surrogate set should span the domain of the species assemblage).

By meeting these properties, insights into the status and trends of the surrogate set allows indirect inference to these same properties for all the unmeasured species (see Noon et al. 2009).

Guidance to selection of the surrogate species set, and identifying the environmental and management factors that drive their dynamics, is aided by developing a conceptual model of the system being monitored (see Manley et al. 2000, Noon 2003 for examples).

In contrast to this approach, the TRPA selected what I would call a "flagship" species approach. That is, they chose to measure high profile species recognized and valued by the public. There is nothing wrong with this approach unless the TRPA is legally mandated, to the extent possible, to sustain all wildlife species within the Basin. Status and trend of many flagship species are unlikely to tell you much beyond their own measurement. A flagship species may be a member of an optimal surrogate set but other species with contrasting ecologies, life histories and spatial and temporal domains would also be needed.

Given the surrogate set, the next decision is what should be measured on these species. Monitoring state variables could include abundance, density, survival rate, or occupancy, for example. The TRPA primarily used count (abundance) data as the state variable. Once the state variables are selected, the sample unit—size, shape and number—for accurate state variable estimation must be described. Collectively, the species selected for measurement, their associated state variables, and sample unit properties constitute the **response design** of the monitoring program.

Next for consideration is the **sample design**. In addition to sample unit properties, a monitoring program needs to specify how the units will be selected for measurement. Typically, in order to make unbiased inference to the target population, sample units must be selected probabilistically so that all possible sample units with the sample frame have a non-zero probability of inclusion within the sample. It is important to remember that statistical inference flows from the analysis (i.e., simple linear regression in this case) to the sample to the sample frame to the target population. All these components must be appropriately integrated.

In my review of these chapters, I was unable to identify the response and sample designs used by the TRPA. It actually seems as if no sample units were designated, at least not in a probabilistic fashion. Reference is made to "population sites" but nowhere did I see this term clearly defined, nor do they seem equivalent to sample units in a traditional survey design. **It is my understanding that the TRPA wildlife monitoring program consists of a single sample site, the entire Lake Tahoe Basin**. Within this single site, over multiple years, monitoring state variables (e.g., number of active osprey nests) are estimated. Inferences to population trend are then made after regressing the state variable on time (year) and determining if the model suggest a significant slope parameter. Below, I will outline several alternative statistical models for analysis of these time series data.

Thresholds

An ecological threshold is the point at which a relatively small change or disturbance in an environmental driver causes a rapid change in vital ecosystem state variable. Thresholds in ecological systems represent strongly non-linear changes in one or more vital state variables given a relatively small, marginal change in an environmental driver (i.e., a natural or anthropogenic process, event or activity). Even though the TRPA uses the term threshold, I believe they are really talking about monitoring trigger points—the value of a monitoring state

variable that is unacceptable and will trigger some sort of management response. Trigger points are an essential component of an environmental monitoring program. However, they must be accompanied by an explicit statement of the management actions to be implemented once the trigger has been "pulled". I found no discussion in the chapters I reviewed of what actions would take place if the threshold target was not met.

Targeted versus Surveillance Monitoring

In order to implement management actions to address unacceptable change in an indicator requires *a priori* knowledge of possible drivers and mechanisms underlying the observed changes. In the absence of this information, managers do not know how to response to observed changes in the indicator. As a result, a period of study is required to gain insights into causation which introduces an unwanted time lag in response to the monitoring information. To avoid this problem requires the simultaneous measurement of the indicator and putative environmental drivers—termed prospective versus retrospective monitoring by Noon (2003) and surveillance versus targeted monitoring by Nichols and Williams (2006). Simultaneous measurement of indicators and drivers (stressors) did not seem to be part of the wildlife program.

Concerns About "Population Sites"

Each of the major indicators (goshawk, osprey, bald eagle, etc.,) is referenced by population sites. I will use active osprey nest sites as an example because this seems to be the most easily understood indicator. It is my understanding that each year all previously active nest sites are visited in year t and if the nest is "active" it contributes a "1" to the count of the **total** number of active nests (the monitoring state variable) in year t. The survey process is continued through the list of all previously active nest sites until a total count of active nests in year t is estimated. The total count for year t (N_t) is the response variable in subsequent regression modeling.

Several issues concern me in this "design". For example, what sampling effort is put into finding new nest locations not observed in previous years? Is there a probability-based design to survey other locations within the Basin that may not have a history of containing nest sites? Further, if a historic nest site is lost, say to development, is that site permanently removed from any subsequent trend analysis samples (a big mistake)? The bottom line is that reliable estimates of trend in the number of active osprey nests requires both the possibility of loss and additions to the most recent list of nests in year t-1.

As an aside: The USFWS, as part of the Bald Eagle Recovery Plan, adopted a survey design exactly tailored to osprey nest monitoring in the Basin. It is called "dual frame sampling" and has been applied to trend analysis for multiple wildlife species (see Haines and Pollock 1998). The two sampling frames are the **list frame** and an **area frame**. The list frame contains information on all previously located active nests. The area frame is used to described the geographical boundaries of the region of interest (i.e., the Tahoe Basin). If this method is of interest to the TRPA, I could provide additional details.

Components of Variance in Monitoring Programs

In order to design efficient monitoring programs with high statistical power, it is necessary to understand the sources of uncertainty in data collection and analysis. Continuing with the osprey nest example, the TRPA wants to be able to understand temporal trends in the ecological processes that give rise to the true number of active nests in year t (let's call the total number of active nest in year t, N_t). However, we know that there are two key sources of uncertainty associated with our estimate of N_t. One is **measurement error**—perhaps the nest was active but we incorrectly concluded it was inactive. The second is **sampling variation**. If the true number of potentially active nest sites in the Basin is unknown, and we sample only a proportion of them, then a hypothetical different sample (of exactly the same size n), would provide a different estimate of N_t. Uncertainties due to measurement error can be reduced in many ways, most commonly by making multiple visit to each potentially active nest to increase the likelihood of observing the true state. However, sampling variation can only be reduced by increasing sampling effort. The goal of a monitoring program is to minimize

bias in the **observation process** so as to accurately estimate the **ecological process** (changes in N_t over time). Stochastic environmental variation directly affecting numbers in count data over time is called **process variance**.

Estimation of Trend—Statistical Issues

TRPA estimated trend by regressing count data (e.g., number of active osprey nests) on time. Let count at time t be N_t (response variable) and year be indexed by t (independent variable). The fitted model was: $\hat{N}_t = \hat{\beta}_0 + \hat{\beta}_1 t$. Note that this model is linear and additive—that is, it models arithmetic population change involving adding or subtracting a fixed amount each year.

However, this not the way biological populations change over time. Assuming no density-dependence, biological populations change by a **fixed rate** each year so that the expected numerical change among years depends on both the **rate** and the **current population size**.

Population trend in most vertebrate wildlife populations is defined as the average change in log abundance per unit time (Dennis et al. 1991). As a result, the appropriate model is **discrete geometric population growth**: $N_{t+1} = N_t \lambda$, where λ = the geometric growth rate. The model for osprey, and for most vertebrate species, is discrete because these species breed once per year, not continuously. Note: λ is easily understood as the logical growth parameter by solving for λ : $\lambda = N_{t+1} / N_t$ Generally, this model is written as: $N_t = N_0 \lambda^t$ where N_0 is the initial population size.

Historically in wildlife ecology, the parameters of this model were estimated by a log-linear regression of counts against time (using simple linear regression, SLR) after transforming the geometric model by the natural log function (ln): $\ln[N_t] = \ln[N_0] + \ln[\lambda]t$. After fitting the SLR model, parameters estimates are interpreted as: $\hat{\beta}_0 = \ln[N_0]$; $\hat{\beta}_1 = \ln[\lambda]$. Estimates of the population rate of change are given by: $\hat{\lambda} = \exp(\hat{\beta}_1)$. If $\lambda > 1$ the population is increasing; if $\lambda < 1$ the population is declining. In general, this is the model traditionally fit to wildlife count data (Eberhardt and Simmons 1992). However, better trend models have recently been developed (which I will discuss, below).

Better ways to estimate population trend from a time series of abundance data are now available (Humbert et al. 2009). Humbert et al. (2009) develop analytical methods to rigorously estimate exponential population trend and its variance from time series of count data thus allowing reliable inference to whether a population is increasing, declining or remaining stationary. Estimation of trend from time series data depends critically on assumptions regarding whether process variance, sampling variation, or both are affecting the variance in growth rates over time (Humbert et al. 2009; Mills 2013).

Exponential growth with observation error: The first model discussed in Humbert et al. (2009) is the discrete geometric population growth model outlined above. When this model is log transformed, the growth rate can be estimated by SLR with $\hat{\lambda} = \exp(\hat{\beta}_1)$. This model assumes that the only source of variance in the time series of counts is due to sampling variation or observation error. The dynamics of the population are assumed to be governed by deterministic exponential growth.

Exponential growth with process variation only: The second model assumes exponential growth with process noise only—that is, no uncertainty exists in the count data due to sampling variation or observation errors. This would be the case if the entire population of interest was censused and observed without error. This model is equivalent to the density-independent diffusion approximation model developed a number of years ago (Dennis et al. 1991, Morris and Doak 2002). This model can be easily implemented in Excel based on a SLR of

transformed time and count data. Specifically, the response variable y_i and independent variable x_i defined in terms of the original data are:

$$y_i = \left(\ln\left[N_{i+1} / N_i\right]\right) / x_i$$
$$x_i = \sqrt{t_{i+1} - t_i}$$

The slope of the linear regression of y_i on x_i, forcing the intercept to be 0, is an estimate of the population growth rate, $\hat{\lambda} = \exp(\hat{\beta}_1)$. Because of the nature of the x_i transformation of the time variable, this model easily accommodates missing data.

Exponential growth with process variation and observation error: For most wildlife species, their abundance estimates vary due to environmental stochasticity (process variation), observation errors and sampling variation. In addition, many abundance time series are subject to autocorrelation due to correlated process variation—that is, a series of good years is often followed by a series of bad years. Autocorrelation in the response variable is thoroughly addressed in this model. This model, referred to as a **state space** model, is fit with a linear model including both random and fixed effects. In my opinion, this model (Humbert et al. 2009) is the most appropriate statistical algorithm to use to estimate trend from the wildlife count data collected in the Tahoe Basin.

Using the number of active osprey nests as an example, I have fit the original TRPA regression model, the observation error model, and the process variance only model using Excel. This spreadsheet is included with my review comments. The preferred trend model incorporating both sampling and process variation, however, cannot be easily programmed in Excel. Fortunately, Humbert et al. (2009) have made available a user friendly R script to fit this model to abundance time series data. This R algorithm estimates growth rates, observation error and process variation for all three models. This script is also attached to my review.

Example Results: Trends in Number of Active Osprey Nests

Summary results of fitting the three models of Humbert et al. (2009) to the time series of the count of active osprey nests over time are shown below (Table 1).

Table 1. Model results: Temporal trend in number of active osprey nests, Lake Tahoe Basin, 1997-2015. LCI = lower 95% confidence interval; UCI = upper 95% confidence interval. σ^2 is the process variance, τ^2 is the observation error.

Model	μ = ln[λ]	σ^2	τ^2	Var(μ)	LCI(μ)	UCI(μ)
Observation Error	0.0288	NA	0.0232	NA	0.0163	0.0413
Process Variance	0.0509	0.0218	NA	NA	-0.0173	0.1191
State Space	0.0509	0.0218	~0.0	0.0012	-0.0173	0.01191

Below, I have plotted the observed and predicted values of the count for the Observation Error Model (Fig. 1) and the Process Variance Model (Fig. 2) to show how well the estimated models fit the observed data. As is clear from Figure 1, the Observation Error Model is a poor fit to the data.

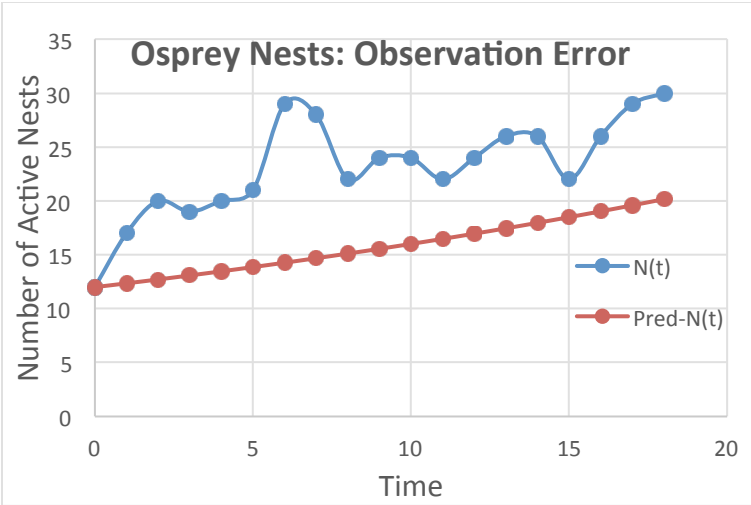


Figure 1: Fit of observation error model to active osprey nest count data.

In contrast, the fit of the process variance model to the data is much better (Figure 2). This is confirmed by the components of variance analysis which shows almost all the variability in the count data to be attributable to process noise and not observation error (Tabled 1). I have not graphed the state space model because the estimate of the growth rate from that model is equal (to four decimal points) to the process variance model.

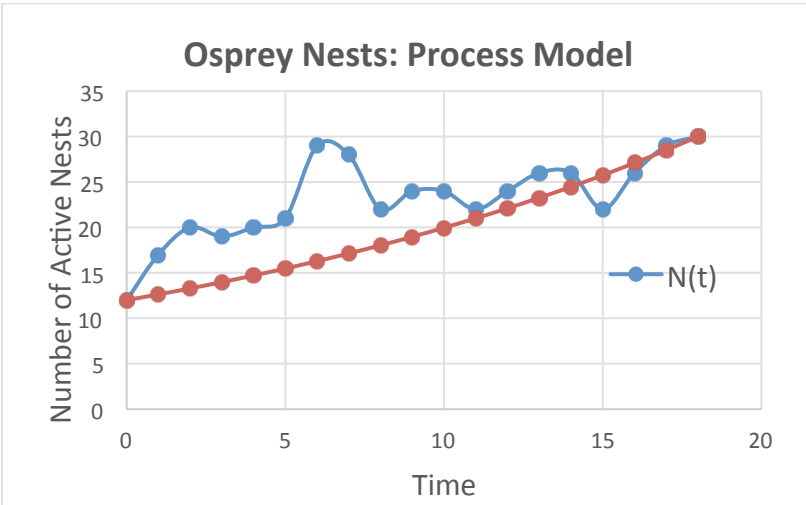


Figure 2: Fit of process variance model to active osprey nest count data.

Based on the point estimate of ln[λ] from the state space model, I would cautiously infer that the number of active osprey nests appears to have increased over time at a rate of approximately 5%/year (= 1 – exp[0.0509] * 100%). I emphasize caution because the confidence interval on ln[λ] includes 0—that is, estimated annual rates of change based on the confidence intervals vary from a decline of 2%/year to an increase of 5%/year.

Brief Comments on Deficiencies Associated with Individual Species

For the most part, the Report (Chapter 8), honestly points survey design and data limitations.

- Goshawk: lack of standardized survey effort or any attempt to adjust the count for variable survey effort; high year-to-year variation in the count undermines reliable estimates of trend unless the length of the time series is long; a probabilistic survey should be designed for this rare and relatively cryptic species.
- Active Osprey Nests: evaluated in detail, above. Dual frame sampling should be considered.
- Winter Bald Eagle Count: methods I used to estimate trend for osprey nests are applicable to these data; how is survey effort standardized across years?; are the counts adjusted for varying survey effort and observer error? consider dual frame sampling.
- Nesting Bald Eagle: data too sparse for trend analysis.
- Golden Eagle: data too sparse for trend analysis.
- Active Peregrine Falcon Nests: methods I used to estimate trend for osprey nests are applicable to these data; consider dual frame sampling.
- Waterfowl Population Sites: This indicator is very poorly defined and its ecological information content appears to be highly uncertain; consider dropping the indicator.
- Disturbance-free Zones: This "indicator" is amenable to trend analysis based on periodic updates derived from remotely sensed data that provide estimates of extent, and spatial pattern (multiple landscape/fragmentation metrics could be considered).
- Riparian Habitats: This "indicator" is amenable to trend analysis based on periodic updates derived from remotely sensed data that provide estimates of extent, and spatial pattern (multiple landscape/fragmentation metrics could be considered).

Literature Cited and Recommended

Dennis, B., et al. 1991. Estimation of growth and extinction parameters for endangered species. Ecological Monographs 61:115-143. 56:603-610.

Eberhardt, L.L., and Simmons, MA. 1992. Assessing rates of increase from trend data. J. Wildlife Management 56:603-610.

Foley, M.M., et al. 2015. Using ecological thresholds to inform resource management: Current options and future possibilities. Frontiers in Marine Science 2:1-12.

Haines, and K.H. Pollock. 1998. Estimating the number of active and successful bald eagle nests: an application of the dual frame method. Environmental and Ecological Statistics. 5: 245-256.

Humbert, J-Y. et al. 2009. A better way to estimate population trends. Oikos 118:1940-1946.

Lande, R., et al. 2003. Stochastic Population Dynamics in Ecology and Conservation. Oxford University Press.

Link, W.A., and J.R. Sauer. 1998. Estimating population change from count data: applications to the North American Breeding Bird Survey. Ecological Applications 8:258-268.

Manley, P.N., et al. 2000. Monitoring ecosystems in the Sierra Nevada: the conceptual model foundation. Environmental Monitoring and Assessment 64:139-152.

Manley, P.N., et al. 2004. Evaluation of a multispecies approach to monitoring species at the ecoregional scale. Ecological Applications 14;296-310.

Mills, L.S. 2013. Conservation of Wildlife Populations: Demography, Genetics, and Management. Wiley-Blackwell.

Morris, W.F., and D.F. Doak. 2002. Quantitative Conservation Biology: Theory and Practice of Population Viability Analysis. Sinauer Associates.

Nichols, J.D., and B.K. Williams. 2006. Monitoring for conservation. Trends in Ecology and Evolution

Noon, B.R. 2003. Conceptual issues in monitoring ecological systems. Pages 27-71 in D.E. Busch and J.C. Trexler, editors. Monitoring Ecosystems: Interdisciplinary approaches for evaluating ecoregional initiatives. Island Press.

Noon, B.R., et al. 2009. Multispecies conservation planning on U.S. Federal lands. Page 51-84 in J.J. Millspaugh and P.R. Thompson, III., editors. Models for Planning Wildlife Conservation in Large Landscapes. Academic Press.

Noon, B.R., et al. 2012. Efficient species-level monitoring at the landscape scale. Conservation Biology 26:432-441.

Review by Dr. Kevin Rose

Rensselaer Polytechnic Institute

Chapter 1: Introduction

Main points:

Overall, this is a well-written high-level overview section. I particularly like the components of the reporting icons of indicator state. They are well described and can translate quite a bit of information quickly. The chapter provides sufficient background necessary to understand the purpose and scope of the Threshold Evaluation Report.

The authors use the introduction to cover the structure of the report rather than provide a high-level summary of the most important findings from the report. Both of these approaches can be useful, but the lack of the summary of the most important overall findings begs the question of it this summary exists and if the authors feel that it should be written. It could be of use to many policymakers and other stakeholders who might not dig into the specifics of each chapter in substantial detail. This sort of summary could look something like the summary for policymakers from the IPCC reports.

Minor points:

Page 1-1: "From 1860 until the turn of the century..." The authors may wish to refer to the turn of which century; I assumed they meant 1900.

Page 1-3: "These policy actions made an enormous contribution to turning around the decline in Lake Tahoe's water quality". It would be helpful if the authors articulated how these policy actions contributed and what water quality response metrics to which they are referring. This is important because in the last paragraph on the same page the authors state that, "...subsequent planning approach were found insufficient to protect Lake Tahoe's irreplaceable ecosystem." This comment addresses the question of whether statements of fact are supported with appropriate references or original data and analysis.

Chapter 2: Methodology

Major comments:

Overall, this chapter is clear and articulate. It is clear the authors have put substantial thought into how thresholds are evaluated. I have a few questions on why certain statistical methods were used, and I think a summary table of all response indicators (thresholds) would be useful, with the table reporting the statistical trend, dataset duration, and any pertinent notes (e.g., data gaps, or depth of samples).

Minor comments:

There is some duplication between the introduction and methods chapters. This is fine, but the authors may wish to consider keeping some of the methodology exclusively in the methodology chapter, such as a description threshold standards.

Page 2-6 to 2-7: The authors state that simple linear regression is the primary analytical tool. This statistical test may not be appropriate for many of the datasets. For example, the test assumes data points are independent, which is not necessarily true in this case. If linear regression is used it would be good to state if there was autocorrelation in the time series data. A more appropriate test that often yields similar slope trends is Sen's slope. On a related topic, in the water quality chapter it seemed like the Mann-Kendall test was used much more

frequently than linear regression. If this test is commonly used in other chapters it might be worth a brief summary in the methods section.

Page 2-7: “Methods for modified analytical approaches are explained in the chapter or individual indicator summaries.” It would be helpful to have a list of all indicators and a column with the statistical test applied to the indicator, as well as the time frame over which the analysis was conducted.

Page 2-9: Currently the text says, “No attainment date is estimated if the threshold standard is determined to be in attainment.” Instead of saying this, it would be more appropriate to say the threshold standard is “already attained” or “currently in attainment”.

Page 2-10: The text says, “Where the data collected meets the “duration of trustworthy data” criterion, and both visual inspection of the plotted data and the test for statistical revealed both indicate that a trend determination...” It would be helpful if the authors articulate what is examined in the visual inspection. This can come across as subjective, but as a reviewer I see high value in visual inspection for things such as non-random residuals. Therefore, it would be helpful if the authors describe what things are visually inspected.

Chapter 4: Water Quality

Overall comments:

There are a wide array of indicators of water quality that are reported, and the chapter presents both a rigorous scientific review of each indicator as well as a more holistic summary of trends and status for each indicator, which can be of value to a broader audience. General weaknesses perceived throughout sections included: (1) making sure appropriate and consistent time-series statistical approaches are used (e.g., Mann-Kendall or Sen’s slope), and (2) considering whether “Insufficient Data to Determine a Trend” is the same as no trend observed. The authors may wish to consider a threshold time period or minimum of recorded samples be used to determine whether there is insufficient data to determine a trend, rather than the presence of a significant long-term trend. Often, no trend is a good thing, rather than an indicator of insufficient data.

In many sections, the authors refer to substances that absorb and scatter light. In general, in organic particles such as suspended sediments from runoff are highly scattering but weakly absorbing. In contrast, organic particles are generally substantially more absorbing and less scattering. So changes in substances such as primary productivity altering organic particules are likely to affect absorbance far more than scattering. The authors may wish to review J.Kirk’s Light and Photosynthesis in Aquatic Ecosystems for a summary. The differential processes of absorbing and scattering will determine how changes in different substances (and substance particle size) influence clarity and color, and the effects of changes in particular response indicators will not necessarily impact both absorbance and scattering equally.

It would be helpful to introduce discuss of limiting nutrients in Lake Tahoe because there is discussion in many sections on the influence of P and N. Not all forms of these nutrients will influence primary production, and it is likely that one nutrient is more important than the other - although this can vary based on time of year and depth, etc.

Early in the manuscript, perhaps in one of the summary tables, it would be useful to describe the statistical test used for each indicator. Sometimes descriptions of the test method is very detailed, while other times (for the same statistical test) the description is absent. Instead of repeating it, it would be useful to have a summary of the tests and why they were selected and to what indicators they were applied. Additionally, for many indicators, trends are assessed using both a statistical test as well as “visual inspection (i.e., qualitative)”. Qualitative inspections should be avoided when at all possible to reduce subjectivity. Finally, some indicators have error bars (typically representing intra-annual standard deviations) while others do not, despite having the

available data to calculate these error bars. It would be nice to see a consistent approach of either including or excluding error bars depending on the type and volume of underlying data.

Many tributary indicators were sampled at base flow, spring snow melt, and storm conditions. For these indicators, it would be useful to see plots of how each indicator varied with respect to flow volume, and perhaps flow volume during these different conditions relationships may vary depending on time of year and previous conditions.

Finally, predictions of future values for response indicators often does not include referencing how response indicators will respond to climate induced changes in air temperature and precipitation, including impacts via changes in stratification and mixing and changes in the timing, duration, and magnitude of precipitation. It would be helpful to include projections, especially when changes in climate are likely to substantially impact the response indicator (e.g., Sahoo et al. 2016). Tahoe is a gem and meeting the various regulations and standards may not in some cases be sufficient in order to improve water quality conditions to targets given non-stationary climate conditions.

Specific comments:

Page 4-1: "Fine suspended sediment and nutrients that support algal growth (nitrogen and phosphorus) are the primary pollutants of concern in the Region because of the negative impact on transparency (Lahontan & NDEP 2010b) and blueness (Watanabe et al. 2016) of the lake...". Also, the header has the text, "Chapter Contributors" but no names are listed.

Watanabe et al. 2016 show that blueness is correlated with chlorophyll fluorescence (an indicator of primary production, and often related to nutrient quantity and quality) but was not related to fine suspended sediments. Therefore, this sentence should be revised, as it implies that fine suspended sediment impacts blueness.

Page 4-1: In this same sentence the authors write, "adversely affected by these pollutants (Reuter et al. 2009)." Are nutrients pollutants? Perhaps this is a philosophical question. I would prefer to see "substances" rather than "pollutants" used.

Page 4-2, line 2, pollutant should be plural.

Page 4-2, the authors suggest that compact town centers well serviced by transit, pedestrian, and bike infrastructure will result in less pollutant load. In theory this would reduce the local N output, but two questions remain: first, earlier in the paragraph the authors state that fine sediment is the pollutant known to impact lake clarity. Are the authors trying to say that compact development will reduce fine sediment flux into the lake? If so, how? Secondly, if the authors are trying to say that compact development will reduce N output and thereby reduce phytoplankton: 1) how much of the N budget comes from local sources? (It is unclear whether the authors are saying that local automobile usage contributes half the total N budget, or if automobile use anywhere contributes half the total N budget) and 2) are the phytoplankton N limited? (Overall, this paragraph needs some clarity - no pun intended :-))

Table 4-1: Is the vertical extinction coefficient what is often referred to as the diffuse attenuation coefficient (Kd) or beam attenuation coefficient, otherwise known as beam-c? If it is the former, the unit of measure description is wrong - scattering not directly contribute to light loss if the light scatters directly back to the surface and out of the system; scattering mostly contributes to increasing photon pathlength, which thereby increases its chance of being absorbed). It would also be helpful to describe what wavelengths are studied. The extinction coefficient is very wavelength dependent and the fact that a minimum of 0.08 is set as the standard suggests that a specific wavelength or waveband is focused on.

Table 4-1: Note that to achieve a turbidity load of <1 NTU, you don't necessarily have to change the sediment load per se. Nephelometric turbidity is a measure of scattering, which is influenced by particle size, shape, and chemistry/particle make up.

It would be helpful to present Figure 4-1 before Table 4-2.

Table 4-2: I don't understand the pie charts for response indicators of "concentration of total phosphorus in tributary waters" and "concentration of total nitrogen in tributary waters". Do the different colors map onto different tributaries?

Table 4-3: The arrows indicate change in each indicator, but change relative to what? The earliest sample?

Page 4-21: Why has winter average Secchi ceased being used as a response metric? It seems like winter average Secchi is where the improvement has been. The analytical approach sounds reasonable except for this piece, "The GAM permits a nonlinear relationship by fitting a smoothing function which allows the trend analysis of recent years to be controlled more by recent measurements." How would a trend of ONLY recent years be controlled by another other than recent measurements? Furthermore, the authors may wish to consider alternative analytical approaches. Specifically the seasonal Sen's slope may be more appropriate to use. The approach doesn't assume independence of sample points and is able to account for seasonality in data, and is frequently applied to time series ecological data (e.g., O'Reily et al., 2015 in GRL).

Page 4-24: In the relevance section it would be good to refer to reader to the Watanabe reference on chlorophyll influencing blueness. Additionally, phytoplankton contribute very little to scattering, so I suggest removing this phrase from this same section. Most scatting is via inorganic particles (e.g., sediments from the watershed).

Page 4-25: It would be helpful to discuss experiments that test for which nutrients are most limiting, and if co-limited, how this varies annually. The report currently states that both nitrogen and phosphorus inputs are increasing primary production - does this mean the lake is co-limited? Also, while Secchi has intra-annual error bars (standard deviations), primary production measurements do not. Why not? Finally, the methods on this page state that simple linear regression was used for analysis. The data do look very linear (impressively linear, really), but least squares linear regression assumes the data points are independent - can you assume that here? Perhaps a more appropriate statistical test would be Sen's slope. Its very likely that the results (e.g., estimated slope, significance, etc) will be essentially the same as linear regression, but the test is more appropriate for time-series.

Page 4-26: "For the period of record (1968 to 2014) phytoplankton PPr has increased by 3.3 percent annually." Is this the average of differences for every pair of consecutive years? "compliments" should be "complements".

Page 4-27: "Interim Target – Reduce the rate of increase in PPr". Annual rate? As the results on the previous page show, the calculated rate depends on the analytical method used.

Page 4-28: "Trend – Insufficient Data to Determine Trend". Does this mean that there is no trend? The dataset looks sufficiently long to determine a trend. How was insufficiency determined? To me it just looks like there isn't a strong trend, especially since the early 1980s. Note: comments above on Table 4-2 on the vertical extinction coefficient also apply here. The wavelengths used as especially useful because the coefficient is wavelength specific.

4-28: "The below average stream inflows and stormwater runoff due to the continuing drought, are substantial contributing factors in the recent improvement of lake transparency (UC Davis - TERC 2015)." Only the last two years look decent, and this is only relative to the preceding year. Is this the "recent improvement"? It doesn't look like an improvement relative to the long term mean.

Page 4-33: Like some other indicators, what does it mean if there is insufficient data to determine a trend? The absence of a trend does not mean that there is insufficient data - and it is possible to collect data for centuries and not have a significant trend. It would be good if the authors describe the minimum duration (consecutive and/or total number of years) needed before a trend is estimated.

Page 4-33 and 4-34: Turbidity (especially when measured via nepholometry) is not the same thing as clarity. While littoral areas are too shallow to use a Secchi disk, there are other methods to estimate clarity. For example, see Gall et al. 2013 in L&O: http://onlinelibrary.wiley.com/doi/10.4319/lo.2013.58.2.0443/full. It would be better to rename this indicator, "Nearshore turbidity".

Page 4-35: According to the methods, "the nearshore extends from the existing shoreline to the average thermocline depth in mid-summer, or a minimum lateral distance of 350 feet from the shoreline, whichever is greater." Does this mean that the nearshore range varies year to year? Or is a particular depth used (e.g., the long term summer mean thermocline depth), in which case it would be useful to just state this value in this section. If the nearshore ranges year to year, you are not comparing measurements over the same spatial area, leading to data interpretation problems.

Page 4-36, and 4-37: Are tables labeled sequentially in this chapter or by indicator?

Page 4-39: Analytic Approach: Based on the results from nearshore turbidity, work needs to be done to decide on the temporal window over which a mean (or peak?) turbidity value will be selected in order to determine compliance with regulation. More work (as stated) is needed on transmissivity, and as suggested above, other approaches to assessing near shore clarity should be considered.

Page 4-45: Some of the species mentioned as introduced unintentionally were likely introduced intentionally, but not by a government agency but rather by individuals. For example, Blue Gill and Large Mouth Bass in the Tahoe Keys area were likely introduced intentionally for fishing.

Page 4-47: Is there "little to no change" or insufficient data in this case? It seems like there is insufficient data to determine a trend here. Sampling hasn't been going on for that long, and it is by necessity species specific. Perhaps it would be better to describe trends in indicator AIS rather than all AIS.

Page 4-48: How do you know that, "The prevention program has successfully prevented the introduction of new AIS into the lake." No new introductions is not necessarily evidence of program efficacy, just as new introductions would also not mean that the program isn't working.

Page 4-52 and 4-53: It would be useful to put a n= to indicate the sample size for each stream.

Page 4-59: The authors may which to consider moving toward a high frequency sensor based approach to assessing suspended sediments, using a nephlometric turbidity sensor. This would remove any potential bias due to sampling dates.

Page 4-61: the pie chart legend is missing the description for the color red.

Page 4-62 and 4-63: the graphs were missing in my pdf version. Only the axis labels showed up.

Page 4-63: The authors state that algae support the food web, but that persistently high algae would be undesirable. This is true, but Tahoe manages to minimize algae/primary productivity and therefore wording in this section might be reworded to avoid the connotation that algae are a "good" thing, given that the management goals are to minimize it. Thus, perhaps it is best to keep this section written in terms of clarity and blueness. Speaking of which, blueness was mentioned early in the report but has not been referred to in any later sections on primary productivity or algae.

Page 4-69 and 4-70: It would be useful to have error bars for the nutrients.

Page 4-72: How closely did the two methods compare? Some statistics on the regression would provide the reader confidence that you can convert measurements of dissolved nitrate plus nitrite between the old and new method.

Page 4-82: Word here is confusing: "Status – Low. Where insufficient data exists to determine status, confidence in the status determination is low. High. The confidence in an estimate of total annual suspended sediment load depends on the number of samples, and on the variance of the daily loads that are sampled to derive the annual load." Also, generally if the p value is very low, you can just write $p<0.01$ or similar.

Page 4-85: Why is the annual used, when the data are sampled at a much higher frequency? And what does the "No" text mean on the graph? This question on data reporting vs. sample frequency also applies to the next couple of response indicators. Plots of precipitation amount versus response load (e.g., suspended sediments, P, or N) would be useful and could be placed in the context of long-term trends in extreme precipitation events. That is, given what the relationship between flow and inputs is, what will future inputs look like given future climate scenarios and trends in extreme precipitation events?

Page 4-92, "The total phosphorus load for each day at each stream was estimated from multiple regression of measured values." What was regressed on what in this instance?

Page 4-121 (and 4-126, 4-131): For just about every other indicator, trends through time are shown as the primary graphic, but not for N concentration. Suggest changing this so it is consistent. Also, in many sections the text, "Because Lake Tahoe is an ultraoligotrophic lake, it is desirable to maintain primary productivity at low levels." is written. Just because a lake is ultraoligotrophic doesn't mean it is desirable to maintain low primary productivity. The justification of why depends on the value associated with different trophic statuses. It would be more accurate to say "Lake Tahoe is an ultraoligotrophic lake and management goals include maintaining this status due to its historic, cultural, economic, and aesthetic value." Or something similar.

Review by Dr. Scott Spak

University of Iowa

Introduction

Considering the Threshold Evaluation Report as a routine planning update, the Introduction provides sufficient background to understand the purpose and scope of the Threshold Evaluation Report. For other purposes and stakeholders, the Introduction lacks an update on recent events or trends that have affected the basin's environment and its human inhabitants, and could contextualize the changes in the indicators studied. The scope is focused by omitting brief consideration of "big picture" changes in the basin.

Statements of fact in the Introduction are generally not supported by references. There are two sources cited other than TRPA and state agency documents. The history section is almost entirely without references. As the report is intended to "summarize and not repeat information already contained in other citable documents," the limited references are of low concern.

Methodology

Prescribed approaches for determining the status and trend of indicators relative to adopted standards are clearly presented and generally appropriate. Assessment of trends over the entire monitoring record or in one or more shorter time series on a case-by-case basis is warranted, but not as codified as the other analytical parameters. "In some cases, with a long- term dataset, the trend arrow depicted in the reporting icon represents examination of the most recent data thereby characterizing the near-term trend." This differentiation is important and valuable. The analysis would provide maximum value to decision support by assessing trends over the adopted TRPA and state planning time horizon(s) for to the projects and plans that address the indicator.

The prescribed approach for determining the level of confidence in status and trend determinations is clearly presented. The selection of r^2 as the primary determinant of confidence in the trend is ideal. Adjusting confidence in trends through t-test p value for such small datasets raises several concerns. The small datasets and high p values used to make determinations are less important than these limitations:

1. <u>First, determine that detrended anomalies in Indicators are random, or quantify observed signals.</u> The analysis assumes that year to year variability not explained by the trend is normally distributed due to observational detection limits, randomness in the underlying data, or both. However, inter-annual variability in most Indicators reflects a combination of management, climate, and socioeconomic signals. Long-term monitoring may be sufficient to support such a multiple linear regression analysis now or in the future after more monitoring at the same sites with identical methods. Results of that analysis would inspire confidence in management effectiveness, quantify the influences of climate and regional socioeconomic trends, and support a comparison of the role of each. Until then, the analysis should technically first confirm that detrended data are normally distributed before applying p value, select an alternative statistic for confidence that applies to non-Gaussian data, or use r^2 alone.
2. <u>Apply alternative approaches when Indictor extremes are not normally distributed.</u> Many of the air quality Indicators quantify the highest hourly or daily concentrations in a one- to three-year period—the extreme highs of the dataset, which rarely follow a normal distribution. In these cases, extreme value theorem provides alternative means for determining confidence, or r^2 alone may suffice.

The approaches prescribed to determine an interim target and attainment date for an indicator are well reasoned, and consider the relevant sources of uncertainty.

Minor edit: the link to Nevada Administrative Code - 445B.22097 is now https://ndep.nv.gov/baqp/monitoring/aaqstd.html

Chapter 3. Air Quality

The air quality section addresses criteria air pollutants for which national and California standards have been long established, supplemented by deposition standards for water quality and management standards that reflect local emissions and emissions proxies. This consistency in regulatory standards and monitoring supports the quality and comprehensiveness of the indicators, monitoring, and analytical methods. As a result, the write-up associated with each indicator evaluation is correspondingly clear and complete, and reflects refinement in the 2006 and 2011 Threshold Reports. While not a formal Indicator in air quality management, the summary analysis of daily AQI is highly informative for trends assessment for a wide range of stakeholders.

Updates

The draft report misses a few recent regulatory and monitoring updates since the 2011 report:

1. The National Ambient Air Quality Standard for annual average PM2.5 was revised in 2012 to 12 $\mu g/m^3$, so the federal, Nevada, and California standard are now the same. The Nevada DEQ (https://ndep.nv.gov/baqp/monitoring/aaqstd.html) and US EPA (https://www.epa.gov/criteria-air-pollutants/naaqs-table) reflect this change. While the California standard is annual, the chart and write-up (page 3-52) might be revised to reflect this synchronization.

2. The Tahoe City FRM PM2.5 monitor conforms to NAAQS reporting requirements. While not yet used in trends analysis, the Monitoring approach for 24-hour (and annual average PM2.5 (page 3-53) should be revised accordingly.

3. Management decisions based on the 2015 report should also consider the role California's 2016 Mobile Source Strategy (http://www.arb.ca.gov/planning/sip/2016sip/2016mobsrc.htm), which is expected to influence many air quality indicators.

Omissions

The report should state that the IMPROVE network does not sample every day, and therefore may not observe the highest 24-hour concentration or accurately resolve the complete annual average on which standards are based. This difference, along with the distance between monitor locations, may also contribute to the higher concentrations observed at Tahoe City.

Analytical Methods

In nearly all cases, the analytical methods are appropriately applied in the determination of an indicator's status, trend, and confidence. A few minor exceptions:

- The NOx total emissions chart includes a trend line from 1975. A 25-year trend from 1990 clearly fits the data more closely, and may be warranted.
- Recent high observations of highest PM10 24-hour average at all monitoring sites (3-43) may challenge the "little to no change" trend determination. A more conservative approach would be "insufficient data to determine" and lower confidence in the trend.

- VMT (3-59) exhibits several trends, and the trend determination is not based on the one long-term trend from 1981 shown. The recent time series employed for trend assessment should be added to the chart.

The 1-hour and 8-hour CO Recommendations for analytical approach request clarification on whether to assess the indicator based on the most current year of data available or on a multi-year average. Considering the discontinuity in monitoring sites, the sensitivity to inter-annual variability in meteorology, and the annual metric in the NAAQS, this reviewer suggests the use of the most recent year rather than a multi-year average.

The 3-year average 4th highest 8-hour ozone Recommendations for analytical approach requests clarification on how to evaluate the indicator. The reviewer supports the reporting of the 3-year average 4th highest 8-hour ozone.

Status and trends analyses

Supplemental datasets, even for Indicator proxies, can help address the limitations of sparse monitors, discontinuities in monitor siting, and differences in observational approaches and frequency between monitors. As the region maintains long-term attainment for each Indicator at state and national monitoring networks, supplemental data becomes less important. The few Indicators with monitoring discontinuities, uncertain or deteriorating trends, or status above threshold would benefit from additional modeling and remote sensing results from routine state and federal agencies to supplement local monitoring. These datasets represent very inexpensive and powerful additions to the information base.

The report notes insufficient monitoring data to assess for sub-regional visibility, an important factor for enjoyment of scenic amenities and currently very close to the TRPA standard on average days. High resolution daily remote sensing retrievals of aerosol optical depth and simulated near-surface extinction and column or surface AOD in regional chemical transport modeling can provide spatial information across the basin that resolve spatial scales (1-4 km) relevant to sub-regional variability and emissions source. The gridded National Emissions Inventory for each primary pollutant and source sector can also serve as an additional proxy. Modeling conducted for regulatory analyses by CARB and US EPA and forecasts from the National Air Quality Forecasting Capability may provide alternative data to support proxy analysis covering the entire basin for every hour. Modeling can simulate Indicator concentrations, trends, and deposition, and attribute them to source sectors.

The visibility Threshold Standard and its discussion should, at minimum, estimate natural visibility conditions (from observations and modeling) and reflect that the US EPA Clean Air Visibility Rule requires Class I visibility areas (including much of the Lake Tahoe Basin) to demonstrate reasonable progress toward natural visibility conditions by 2065. It is also recommended that the agency amend the regional visibility Threshold Standard by including the baseline and adding natural visibility conditions and target trend rates to support the use of the Indicator as a management tool coincident the federal Clean Air Visibility Rule.

Data Support

Statements of fact and conclusions are supported by appropriate references and generally by original data and analysis. Air quality in the Lake Tahoe basin has been studied extensively by CARB and its grantees in recent years (http://www.arb.ca.gov/research/research-results.php). It is recommended that the Threshold Indicator Report update the information base with key applied findings about air pollution in the basin from peer-reviewed and regulatory science studies since the last report.

Minor edits

The 1-hour and 8-hour CO Status discussions (3-13 and 3-16) might report the 2nd and 3rd highest observations from 2012 at the discontinued monitoring site for context in support of the decision to that they were due to a false/faulty reading.

The NOx concentration chart (3-31) should also include the federal standards.

The Nitrate Deposition figure in the 2011 report is contextualized by precipitation data, and its inclusion would help reinforce the 2015 analysis.

Global comment: locations of monitoring sites are difficult to identify on indicator status maps. Larger symbols, please.

Overall Report

In the opinion of this reviewer, the report contains observations of (nearly) all of the environmental indicators ***necessary*** to make informed decisions about environmental management in the Lake Tahoe Basin. The report also clearly and comprehensively addresses uncertainties and spatial and temporal representativeness of the observations made in providing complete documentation for those indicators. As a traditional environmental planning assessment, the report provides the comprehensive management "dashboard" and "rear-view mirror" required to inform decisions. The agency is commended for this clear and extensive approach to indicators of local environmental change, at the leading edge of a regional planning indicator set relative to best practices. The large set of carefully selected, monitored, and analyzed indicators is both commensurate with and tailored to the unique and uniquely stringent environmental management and protection needs of the basin.

From a contemporary environmental planning perspective, environmental observations alone are not entirely ***sufficient*** as "the information base to inform modification of the regional regulatory framework, alter mitigation requirements for proposed projects, and target projects to improve the environmental health of the basin." In the car analogy, there's a very limited "windshield view" for each Threshold Indicator. Several additional information bases are required for rational planning to inform decisions on changes to regulations and new projects:

<u>1. Similar indicators assessing the drivers that cause local environmental changes in these indicators between assessment reports and over time scales relevant to regional planning are required to understand why indicators have changed beyond the expected response to local management.</u>

The most obvious of these are changes to the basin's climate and socioeconomic activities:

a) Local climate in the basin and Sierra Nevada range—especially temperature, snowfall, snowpack depth, and precipitation frequency/magnitude/timing—has a direct influence on most indicators eight of the nine threshold categories: water quality, soil conservation, air quality, vegetation preservation, wildlife, fisheries recreation, scenic resources, and recreational opportunities. These influences are responsible for much of the inter-annual variability in the indicators that isn't explained by the quantified trends. While some of these influences are *described* in the current report, they are not *quantified* as distinct indicators, and the planning process would benefit from their inclusion. From an ontological perspective, many of these climate signals also have obvious "thresholds" relevant to the indicators they impact, and could thus be incorporated consistent with the existing framework.

b) Now that land use/land cover change and resource extraction have slowed dramatically, changes to the region's economy are driven primarily by exogenous forces: skiing/gaming/recreational activities elsewhere in the region and in the world relative to those in the basin; skiing conditions during the season; and the regional/national/international trends in demographics, employment, transportation costs, and disposable income that lead to local visits, in/out migration and corresponding changes to local demographics, and management indicators like VMT. Again, these are currently partially described, but not quantified as indicators. Understanding trends in these drivers and corresponding local socioeconomic changes, even if anecdotally obvious, are essential for management that balances and affirms both environmental health and the local

economy. High quality indicators for these external drivers of local change are readily available from state and federal agencies and TRPA partners without additional investment. Indicators for local skiing and tourism relative competitors are likely available, too.

2. Process knowledge that connects indicator status and trends to driving forces.

It is recommended that the Threshold Indicator Report update the information base with key applied findings about air pollution in the basin from peer-reviewed and regulatory studies since the last report.

Limitations in indicators monitored

One issue with planning by indicators is that the planning scope evolves to primarily consider monitored changes, and responds to unmonitored issues only when they cross a threshold of visibility and impact that demands an immediate triage response. From a SWOT perspective, Threshold Indicators designed around monitoring Strengths and Weaknesses miss the unexpected Threats.

In the case of the Tahoe basin, there are a few important issues that have been studied recently and may eventually warrant preliminary assessment or formal Threshold Indicators:

1. Invasive species and vector-borne diseases impact the integrity of the ecosystem and can impact fisheries, wildlife, habitat, and soil and vegetation conservation directly, with indirect impacts on scenic amenities, water quality, and air quality through downstream effects from forest fires and other ecosystem disturbances. Since the impact threshold is the mere presence of these species, management indicators would be an effective start. As with existing indicators, invasives and their impacts can increase due to climate change.
2. Microplastics and persistent organic pollutants bioaccumulate and can magnify in the food web, impacting fisheries and wildlife for hundreds of years. The lake's long residence time makes it a closed system for them on planning time horizons. Neither are effectively controlled by wastewater treatment, and both can deposit in the ecosystem from local use: microplastics from synthetic fleece clothing, personal care products, and litter; POPs from volatilization and spills. POPs can also arrive through atmospheric transport and deposition. Plastic fibers have also been confirmed to concentrate POPs from water, and POPs can have synergistic effects, especially in endocrine disruption. Studies have confirmed the presence of microplastics in Lake Tahoe and POPs, including polychlorinated biphenyls, dioxins, and polybrominated flame retardants, in its fish. Since these pollutants cannot be remediated as quickly or easily as the criteria air and water pollutants for which indicators have been established, a precautionary approach may be warranted. For these pollutants, monitoring is often more expensive than proactive management, so pairing management indicators with monitoring once per reporting cycle may be an effective approach, with prior studies serving as the first monitoring results.

Review by Dr. Carol Wessman

University of Colorado, Boulder

Chapter 1 Introduction

The introduction provides sufficient background information for understanding the purpose and scope of the Threshold Evaluation Report. The history is useful and gives important management background and perspectives. It appears to be supported well, with appropriate references and original data. I have a few comments and some minor edits.

- It's useful to see the statement regarding the insufficiency of regulation alone, and that guided growth through wiser resource management and infrastructure were eventually emphasized in the Tahoe watershed. I especially appreciate that TRPA recognizes the importance of scale for effective management of environmental and socioeconomic conditions.
- The term "threshold" is fundamental to the evaluation of Lake Tahoe watershed condition and yet I have never seen it well defined in any document. In the ecological literature, threshold means a point of dramatic change. This is not necessarily a desired thing. However, this Threshold Evaluation Report speaks to "progress toward attainment of adopted environmental threshold carrying capacities …". I find this usage of "threshold carrying capacities" confusing. As this terminology is embedded in the whole evaluation process, I think it's very important to be much more clear in your definition of "threshold" and "threshold carrying capacities".

I acknowledge that "Threshold standards" are defined on page 1-5, and this helps in understanding the usage of "threshold". However, the relationship between the "threshold standard" and "threshold carrying capacity" is not clear and needs to be clarified.

<u>Minor edits</u>

- "Comstock-era mining" : should have hyphen between Comstock and era.
- Pg 1-3, second paragraph, second line: "scientific knowledge about the lake and **its** surrounding ecosystem …"
- Pg 1-4, 1st line under "2015 Threshold Evaluation Report" – "This report presents the results of TRPA's sixth comprehensive review of progress <u>against</u> threshold standards…."
  - "against" is not the best word; better would be "based on" --> "progress based on threshold standards"
- Pg 1-6, Under "Numerical Standard". Phrase "…shall not be decreased below 29.7 meters." Can this be reworded? Perhaps "reduced" instead of decreased? Confusing.

Chapter 2 Methodology

The methods for determining the status and trend of indicators relative to adopted standards are clearly presented and sufficient for the data at hand. The determination of confidence levels Is well described, and continuous attention to this process is very important to maintain and/or improve monitoring procedures. Sources of uncertainty should be addressed in the context of confidence levels, but are not discussed in this chapter. Determination of interim targets and estimation of attainment dates are adequately reviewed and appear appropriate for all indicators.

Minor edits

- Page 2-1, last paragraph under "Indicator". "…has a direct relationship to the status of attainment or maintenance of one or more threshold or standards." "Threshold standard" is typically used together in this report. Why the separation here with "or"?
- Page 2-3, the link for Nevada (http://www.leg.state.nv.us/nac/nac-445b.html) is missing.
- Page 2-7, Figure 2.2 is redundant with information in Figure 2.3. Not necessary.
- Page 2-10, paragraph under Table 2-3. Text is missing: "…and both visual inspection of the plotted data and the test for statistical revealed …"

Chapter 5 Soil Conservation

The soil conservation threshold is determined using two indicator reporting categories: impervious cover and stream environment zones. The descriptions of these categories are complete and clearly written. With the exception of a few points of clarification (see minor edits below) the approaches to determine impervious cover are appropriate and useful to track status and trends. The update from Bailey's 1974 classification to the more recent 2007 NRCS surveys is described well and appropriately applied to determine status, trend and confidence. Similarly, the SEZ indicator category is detailed well and clearly argued as a useful indicator of the condition and functionality of stream environments. However, the text describing the Indicator State #4 (Attain a 5% total increase in the area of naturally functioning SEZ lands) is difficult to follow and should be clarified; I.e. Should the target of 5% increase be determined from the overall acres of SEZ in the region, or be considered an increase to the acres of functioning SEZ at the start of the monitoring period?

The discussion concerning the three factors complicating objective and consistent evaluations of status are useful and address recommendations well. For example, the *ambiguous objectives* section suggests ambiguity in terms such as "disturbed" and "naturally functioning" may be vegetation-type specific and require different metrics to characterize the functioning of the SEZ. Absolutely. Under the 3rd factor (*"Presence of multiple clauses*…"), the Roby 2015 SEZ review project is introduced. Its estimates of substantially higher SEZ area in the basin (29,391 versus 21,944 in the Tahoe 2001 plan) should give rise to a serious comparison and revisit of field delineation protocols. Why are these estimates so different? Can they (or should they?) be "calibrated" to one another to keep the valuable time series? It is important that the standards for delineating SEZ are the best available, given the reliance on change-of-area is an indicator of condition.

Minor edits

- "Comstock era logging" should have a hyphen --> Comstock-era
- Pg 5-1, 4th para. "Excessive impervious surface contributes …and its tributaries **by** impairing water quality, …"
- Pg 5-4, 1st para. "…contributes to sediment and nutrient inputs to Lake Tahoe and its tributaries **by** imparting water quality, alter surface hydrology…" *Or use appropriate punctuation.*
- Pg 5-6, 1st para. Use of "nearly doubled" for the change in resolution from 2007 to 1974 is confusing, and could be considered as the 2007 resolution between twice as coarse as 1974. Needs rewording.
- Pg 5-6, 2nd para. This paragraph needs a concluding sentence that refers back to the 2007 classification. The goal of the paragraph is to distinguish the 2007 classification from the 1974 Bailey classification, and half of the paragraph describes the Bailey's approach. Conclude with a statement on how the 2007 handled these issues.

- Table 5-3 caption. "Land capability class 1b ….by 657 acres." Should be 659.6 acres.
- "All other land capability data classes indicate that current cover is above the allowable level." Should be *below* allowable cover.
- Pg 5-17, 1st para, 4th line. "….commonly defined, preservation **is** interpreted to mean…."
- Pg 15-17, 2nd para, 16th line. "….of a strict location criteria where restoration of 25 percent of the SEZ must occur." Delete "*for*".
- Pg 5-20, under *Ambiguous objectives.* "Prior to *the* 2006 report, no location-based criteria *were* used in defining the set of SEZ …."
- Pg 5-21, under *Presence of multiple clauses with the standard*. "Numerous clauses in the standard make interpretation consistent and objective evaluation of the standard challenging." Delete *and*.

Chapter 6 Vegetation Preservation

The long view of this planning process is greatly appreciated and will be valuable forthe preservation of the ecological health and integrity of the Lake Tahoe Region.

Overall, I find the report to be well written and each indicator evaluation clear and complete. Generally, the analytical methods are appropriately applied to determine status, trend and confidence of the indicators. It may not be in the scope of this evaluation, but more incorporation of landscape metrics in the analyses would quantify aspects of pattern (e.g. patch size, connectivity) that would be very valuable in determining indicator status or trends. In their current form, the TRPA evaluation approaches will capture coarse changes in vegetation presence and abundance, but will not always be able to identify critical thresholds in pattern that may lead to decline (or suggest improvement) in certain vegetation types.

Introduction

The introduction to this chapter is well written and presents a useful foundation for diversity of ecosystems in this region. The history is useful for the context of the indicator categories and definitions are usually provided for less commonly-used terms.

- Table 6-1: Why isn't the Subalpine Forest considered in the Relative Abundance standard? All other forests are listed, as well as most other vegetation types. I understand that the 6 indicators, described further down on pg 6-10, are community richness and the 5 veg types identified under Relative Abundance. But a better link between Table 6-1 and the text under Common Vegetation is needed to make that connection. A relevant note here: Text throughout the report could be referencing tables and figures much more (and thus requiring they be numbered).
- I would find a discussion/overview on the concept of the major evaluation intervals useful upfront, in the introduction. Is that possible? It is defined on page 6-14, but it's an important aspect of this whole process. Each category brings in the monitoring intervals, and often in not-so-clear terms. An upfront definition could include either a general table with individualized tables in each of the category evaluations, or clearer presentation in each category, such as a table.
- A discussion of disturbance would be useful in the introduction. This evaluation process equates disturbance with impervious surfaces. That definition needs to be upfront. A more refined definition would include acreage due to other disturbance types, including land use (not just impervious), fire, beetle kill, and fragmentation (e.g. density of trails; patch size), but that is, understandably, a more time-consuming analysis. Ultimately, these analyses are encouraged.
- Table 6-2. The "I" symbol is not defined in Fig. 6-1, and I found no definition until pg 6-32.

Common Vegetation

- Page 6-10, 4th paragraph. Include the time period in the discussion of the prolonged drought. When did it start; i.e. to present, how many years is "prolonged"? When did the bark beetle epidemic start, how long has it been active?

Common Vegetation: Vegetation Community Richness

- Page 6-14. Modification of the Threshold Standard or Indicator
    - A four-year evaluation period is too short to assess trend by itself. This is the sampling interval for long-term monitoring. Make that distinction clear. Are you saying that evaluation every 4 years (assuming in perpetuity) is too long? I would suggest that that is not too long, given the time frame of human activity, and if you are to capture response to climate change, it is as likely to be a drawn out decline as it is to be a "rapid" failure.

Common Vegetation: Relative Abundance of Meadows and Wetland

- Page 6-17. Effectiveness of Programs and Actions
    - In discussion of natural succession wetland and meadows – don't use the term "converted" ("converted to a forest type"). While I see its usage in context, it still brings in the sense of human activity. Use "succeeded to" to keep consistent with the natural process.

Common Vegetation: Relative Abundance of Deciduous Riparian Vegetation Type

- Page 6-21. Monitoring and Analysis
    - Under Analytic Approach – I gather from this text and in other places in the document, disturbance is defined as impervious surfaces. This should be defined upfront, perhaps in the introduction.

Common Vegetation: Relative Abundance of Shrub Vegetation Type.

- Page 6-26. Recommendations: Modification of the Threshold Standard or Indicator.
    - The target for land occupied by shrub vegetation is less than or equal to 25%. I agree that this may be problematic – should a minimum be set for attainment or is no shrub cover a reasonable target in some regions?
    - The third point (top of page 6-27) states that a "simple accounting of the spatial extent" does not provide managers with an understanding of the relative condition of this vegetation type. This is true for percent cover of any of the vegetation types. However it is the first step in landscape assessment, and is a very useful metric in conjunction with other landscape metrics.

Common Vegetation: Relative Abundance of Yellow Pine and Red Fir Forest in Seral Stages other than Mature

- Pg 6-28. The 2001 and 2006 data (<20" dbh) are not comparable to the 2011 and 2015 data (<10.9 dbh). The color schemes (i.e. yellow – yellow pine; red – red fir) suggest that all years are comparable. These should be two separate graphs, or distinguished in some way from one another.
- Pg 6-30. Indicator State
    - Status. Immature red fir forest covers 3.3%, not 3.9%, of the Region.

Common Vegetation (Pattern): Juxtaposition of Vegetation Communities and Age Class

- First place that the Status: Implemented is used (and defined)
- Pg 6-33. Adopted Standards. "Adjacent openings" – are these actual openings in the canopy or would it be more descriptive to call them adjacent patches?

Common Vegetation: Consistency with Bailey Land Capability System

- Line 2. "….flow regulation…" Unclear. Are you referring to water flow? Perhaps hydrology?

Common Vegetation: Non-degradation of Stream Environment Zones

- Monitoring and Analysis. Having the Analytical Approach in the form of a question is inconsistent with the rest of the indicators. The approach should be defined by how the question is answered.

Uncommon Plant Communities

- Pg 6-56. Remote sensing is mentioned as a cost-effective way to assess trends in plant communities. I agree. Remote sensing could be used to monitor greenness, connectivity, and interannual trends in these traits. Some inference of community composition based on patterns might be possible.
- Pg 6-77. Point 3 is misplaced (cut and pasted from previous Indicator). This paragraph is not customized to Grass Lake.
- Pg 6-80. Human & Environmental Drivers, line 13. "….species likely to be replaced by species with wider ecological amplitude." What do you mean by amplitude?

Sensitive Plants

- I don't understand the last sentence on this page: "Galena Creek rockcress, determined to be considerably worse than target, is unlikely to be attainable because the target number of populations have been observed in the Region." The number of populations have reached the target, but are considered worse than the target?
- Page 6-93. I find the text under Monitoring Approach (Draba asterophora var. asterophora) to be very confusing. I can't figure out how many subpopulation sites we've ended up with, and what the monitoring protocol is for those populations.
- Page 6-94. *Indicator State, Status*. Unclear language. First, 4 main populations in LTBMU in 3 locations, at which 41 subpopulation sites are monitored. Then it's mentioned that demographic trends and drivers are monitored in 5 populations with 10 subpopulation sites with permanent plots in the LTBMU and the Humboldt-Yoiyabe National Forest. How do these two sets of populations relate? (4 vs 5? 4 + 5? 4 of 5??) Is an additional one in the Humboldt-Yoiyabe NF? How are "monitoring" at the 4 and "demographic trends and drivers" at the 5 sites different??
    - And *then*, under *Trends*, there is mention of 34 population sites – as those part of the Putnam PhD dissertation – are these overlapped with the 41 subpopulations mentioned under *Status*?
    - Please rewrite these sections for clarity.
- *Long-Petaled Lewisia*
    - Page 6-97, *Analytic Approach*, "….as the maintenance of two population sites can be demonstrated…." Why say two populations here, when your monitoring includes more? Are you implying "at least" two?
    - Page 6-98, 3rd line, spell out SWE à "snow water equivalent"
- *Cup Lake Draba*
    - Page 6-101, *Monitoring Approach*. Start this paragraph with the total number of monitored subpopulations (10), then give the history. Try not to bury the important numbers.
    - "Monitoring occurred every two years …" Is this monitoring different than the census of every 5 years by LTBMU staff? (6th line)
    - *Status*, 1st line, "11 subpopulation sites in the Region" – But only 10 are monitored? (*Monitoring Approach*) What's the difference?
- Page 6-104
    - 11th line, "…. with 41 element occurrences reported …" What are element occurrences?
    - 15th line, there is a reference to the "figure, part a". Here it would be useful to have a figure number.

Late Seral and Old Growth Ecosystems

- Bar graph (again, it would be useful to have figures numbered!!) à the legend needs to define the other two bars; only the 2006 and 2011 estimates are labeled.
- Page 6-112. *Modification of the Threshold Standard or Indicator*, last line. "…. standard for that zone should be considered to more accurately reflect the mature state of species occurring in that zone." Be more transparent with the language. Here "mature state" isn't clarifying the question. Use "age" of species.
- The discussion around old growth is not clear, and the definitions by LTBMU given to early, mid and late seral stages as ranges of quadratic mean diameters will not be consistent across zones. Given that growing conditions at each of the elevational zones are different, size class for old growth will vary. In other words, it's inappropriate and inaccurate to use the same size class for old growth across elevation zones.

Minor edits (clarity and grammar)

- The term "data" is plural. Its usage is mixed between plural and singular throughout the report. Despite popular usage in the singular, this is a scientific document and data should be used in the plural.
- In several places in the document, temperatures are spelled out (e.g. ninedegree Fahrenheit; a two- to four-degree Fahrenheit). This is rather annoying. Using single digit numbers and the degree symbol is fine and far more readable. (e.g. 9˚ Fahrenheit; a 2˚ to 4˚ Fahrenheit, or even 9˚F and 2˚ to 4˚F)
- Page 6-2
    - Line 5. Add "was". "Sheep grazing was ubiquitous in the Region's forests and shrublands, and was so intensive that the understory was often denuded…"
    - Line 8. Please clarify sentence beginning with "A grazing allotment system was put in place …." Is this saying one permittee at a time is allowed for certain areas?
    - 2nd para, 3rd line: "managers" to "manages" "…today the Forest Service manages 78 percent of the Region."
    - 4th para, 1st line: Add "the" à "…poses a threat to the integrity of the Region's vegetation communities…."
- Page 6-10
    - 3rd para, line 10. "…these communities and encourage management practices that promote healthy forests.
- Page 6-13
    - Under Analytic Approach: "Proportion of the Region covered by individual vegetation types is calculated by dividing the area of the vegetation type …."
- Page 6-16
    - Under Indicator State: "Based on this target, the Region is at 93% of the objective of the management target."
- Page 6-33
    - Under Human & Environmental Drivers
        - 2nd line. Delete "A.H." in citation of Nagel and Taylor.
        - 7th line. Hyphenate "…. Fire suppression-focused forest management…."
- Page 6-37
    - Under Human & Environmental Drivers. 1st line, "The structure and distribution of vegetation in the Tahoe Region is in influenced by …."
    - Under Effectiveness of Programs and Actions. 5th line, "….soil conservation chapter of this evaluation which found that more than 10 acres of impervious ….
- Page 6-45

    - 4th para, 3rd line, "…. Which affects their ability to function properly…" (not "property")
- Page 6-49
    - Indicator State: Trend
        - line 5, "…. The spatial extent of community decline suggests…."
- Page 6-54
    - Monitoring Approach, 2nd line. "…. Assess the effectiveness of restoration work, …."
- Page 6-55
    - Interim Target It is not possible to set a numerical target until additional monitoring data are available to gauge the status and trend of the site.
- Page 6-56
    - Attain or Maintain Threshold – "The Conservancy-led project will restore….
- Page 6-60
    - Programs and Actions, 5th line, "… designed to prevent new invasives from being introduced to the Lake …."
- Page 6-63
    - Adopted Standards, last line, "….Freel Peak Cushion Plant Community, …."
- Page 6-65
    - 1st line, "TRPA and partners implement regulations …
- Page 6-73
    - "3. Long-term monitoring of Sphagnum spp and …." First sentence is a phrase and is inconsistent with points 1 and 2.
- Page 6-83
    - Line 1, "Consideration should be given to the use of a mixed effects model to …."
    - Modification of the Threshold Standard or Indicator, line 6, "Standard revision should also give consideration of the likely impacts of …"
- Page 6-87
    - Indicator State. Status. Line 4, "Figure 2 of this indicator sheet." No figures are numbered. As I mentioned above, numbered figures and tables, and more reference to them in the text is desirable to help in the interpretation of the data.
- Page 6-92
    - Under the Trend figure, what is the sentence "The observed increases in population two are likely due to …" referring to? The Heavenly Ski Resort? Please refer to by name, not population number.
- Page 6-98
    - Analytic Approach, 4th line, "… the following analysis to assess trends …"
- Page 6-109
    - Human & Environmental Drivers, 3rd line. Use consistent format on citations. No first names, initials.
- Page 6-110
    - Analytic Approach, line 6. Incomplete sentence? "…. based on (Keyser and Dixon 2015)." Based on what? On Keyser and Dixon (2015).
    - Line 8, "…. dbh, the size of "old growth" forest."
- Page 6-111
    - Programs and Actions, 2nd to last line, "…. Projects are expect to enhance…."

Additional Review by Dr. Sonia A. Hall

Conservation Science Partners

In addition to coordinating the review of each chapter by at least two reviewers with specific expertise in the topic each chapter deals with, I am providing comments on most chapters. It is important to note that my expertise in the chapter's topic is general rather than specific, so these comments should be taken as subservient and complementary to those comments provided by the topic-expert reviewers.

Chapter 3 – Air Quality

P 3-1: Achievement of air quality threshold standards *suggests* – rather than provides partial evidence for – effectiveness of the TRPA Regional Plan, associated programs, and air quality regulations and programs. Actual causal analysis would be necessary (and even recommended) to provide actual evidence of this effectiveness, and lack of notable shifts in indicator values when certain actions were implemented suggests that other drivers may be overwhelming such effectiveness in some cases.

In the Fisheries chapter I highlighted the value of including TRPA's goal in the introductory paragraphs. What is TRPA's goal for air quality? Is it compliance with federal and state standards? Adoption of more stringent thresholds for some of the indicators suggests it is more than that. The goal should be explicitly stated.

Regional Air Quality Conditions as Measured by the EPA Air Quality Index: This section provides useful information, as identified by Mr. Hunt in his review. It is also a good example of ancillary data and analyses that are valuable and informative to decision-makers that rely on these Threshold Evaluation Reports. There is much opportunity to strengthen such ancillary data and analyses, as Dr. Spak describes, throughout the Report. However, this particular section is confusing in this particular place. I would recommend shifting this section (and the reference to the EPA AQI on P 3-8) to the end of the introductory section of this chapter (after Table 3-3 and initial description of the indicators and their status and trends summary). I would further recommend some discussion of how to best use this information to complement the Thresholds Evaluation.

P 3-11: Add to Figure 3-1 a simplified version of Table 1-1, the management and policy reporting icons.

P 3-25: Please describe why data do not meet regulatory requirements for completeness and validation. The implication of this statement is very different if not meeting such requirements is due to (a) data too new and not yet evaluated, (b) not evaluated, and will not be evaluated, or (c) evaluated and found lacking in quality.

Overall: This chapter would benefit from a Conclusions section at the end, that summarizes the status and trends for the whole threshold category. This would be a narrative discussing Table 3-3, which was a very useful, visual summary of all the indicators and their status and trends. This concluding section would address: what do the status and trends information on all these indicators say about the status and trends of Air Quality as a whole? And what are the overall recommendations for the Air Quality category?

There are many air quality indicators that are influenced by prescribed burns and wildfire, and prescribed burning regulatory controls are identified as drivers and/or actions implemented to improve conditions. In Washington State (where I have greater experience) there is an active debate on the trade-offs between strict regulations of prescribed burning to maintain air quality standards in affected communities, and the limits these impose on agencies' ability to restore forests, with consequences for wildfire occurrence and associated, more severe air quality issues in the summer. TRPA should consider supporting and/or funding efforts to understand such trade-offs in the Basin, or discussing them in this Report if they are already fully (or partially) understood.

As discussed by Dr. Spak in his review, there are many opportunities, and a distinct need, to provide additional information, data, and analyses that can inform whether actions are leading to changes in indicators. This is particularly the case where there is a clear hypothesis articulated, and data already exists that would allow some evaluation of that hypothesis.

P 3-35: The distinction between "regional" and "sub-regional" visibility indicators was initially confusing. Please explain where this is first articulated why one monitoring site can provide a "regional" measure, while another site only provides a "sub-regional" measure. If both are individual sites, why do they provide information at different geographic scales?

P 3-38 (and beyond): Applying the same criteria for trend confidence to a trend of "little or no change" as are applied to "improvement" or "decline" trends is counter-intuitive. If the regression analysis suggests there is no trend, then of course the r2 will be low and the p-value will be high. These metrics reinforce the lack of trend, rather than undermining the confidence in the result that there is no change. I recommend consulting with a statistician to evaluate how the r2 and p-values should be interpreted to inform the confidence measures.

The fact that TRPA has no ability to regulate or otherwise control factors such as wildfire outside of the basin does not necessarily mean TRPA is unable to affect those factors. The entities TRPA partners with on monitoring, its ability to inform or fund research on the impacts of those outside factors on Basin values, as well as TRPA's relationship with regulatory agencies that can control some of those factors may be critical avenues for TRPA to influence what happens outside the Basin. As Dr. Spak identified, looking at that "bigger picture" may become increasingly important for TRPA to do, to maintain attainment of some of these indicators in the future.

Is the frequency or occurrence of low visibility days not an important indicator, in terms of the frequency that such air quality concerns affect the scenic values of the Basin?

P 3-47: The connection between indicators is under-used and under-discussed. As an example, knowing (or at least providing hypotheses, to then be tested) why annual average PM10 values are considerably better than target while peak 24-hr concentrations are somewhat worse than target could well provide guidance as to where, when or why current programs and actions are insufficient for bringing the peak values into attainment.

P 3-52: Description is needed on how the annual mean concentrations are weighted to provide the 3-year average.

P 3-60: The assumption that the VMT constant – which depends on factors such as the road network, residential housing units, income, occupancy, employment, etc – does not change across a 25-year window undermines the value of the trend. Cannot the TransCAD model be used to evaluate the sensitivity of that constant to changing assumptions for these different factors? Conclusions from a sensitivity analysis that at least quantify whether the values tend to be over- or under-estimated by considering it a constant would be very informative for interpreting this trend.

P 3-61: Given the statement under Confidence that VMT was estimated with progressively more sophisticated – and hopefully, more precise – models, the statement under Effectiveness that current programs and policies are mostly effective in reducing VMT is unsupported. Isn't it possible that the decreasing trend in VMT could simply be due to increasingly accurate or precise measures, rather than an actual improvement in the indicator?

P 3-63: Why are Human and Environmental Drivers for Odor not applicable?

P 3-66: I support the recommendation for modifying this indicator, shifting from a policy statement to a numerical standard. I further recommend that this should be the approach that TRPA strives for in relation to all its management and policy statements. As information improves and/or opportunities arise to quantify these statements, TRPA should move to do so.

Minor comments

P 3-2: Were the AQI values reported in the last paragraph calculated by TRPA based on EPA data, or provided directly by EPA?

P 3-7: The link to the TRPA Threshold Standards (goals document) is wrong.

P 3-8: Please briefly describe when and why annual summaries versus raw data were needed.

P 3-9: The first paragraph on addressing data limitations should be integrated into the 2nd paragraph in the Data Limitations section on P 3-8.

Confidence in the icons of Table 3-3 are not distinguishable. Rather than shrinking the icon, shrink the size of the icon while leaving the formatting of the border unchanged.

P 3-12: Provide the regression line equation, r2 and p-value for all trend lines in this and other graphs. Use consistent representation throughout.

Cross-references to later sections where a point is discussed in further detail would avoid questions and concerns on the part of the reader. For example, the 2012 CO false reading statement on this page is unjustified. A simple pointer to see the section in the summary where this is discussed would avoid this being an issue.

P 3-17: Who is Campbell? S/he provides critical "personal communications", yet no information is provided that would lend credibility to such information.

P 3-24: The suggestion of alternative definitions that could be used was very useful, particularly in light of Mr. Hunt's specific recommendations on which he found most reasonable. Note however that the alternatives identified on P 3-30 should be edited to total nitrogen oxide emissions.

P 3-28: In multiple trend graphs the legend includes symbols not plotted in that particular graph (e.g. in this graph, the open symbols are not plotted). Please "clean up" the legends to reflect those datasets plotted in that particular graph.

P 3-36: Suggest not linking data points, as it obscures the trend lines, that are more important.

P 3-56: Check the r2, as r2 should not be a negative value.

P 3-60: Type of standard is considered a visibility indicator, yet is not included under visibility in Table 3-3. Why?

P 3-66: Provide references for where the quoted text was copied from.

Missing reference: Lahontan Water Quality Control Board 2010 (also referenced as Lahontan 2010 on P 3-55 – be consistent on formatting of references).

Chapter 5 – Soil Conservation

P 5-1: Is the goal to "protect the Region's soil resources and provide their continued ability to filter and retain nutrients for a variety of purposes"? Articulating clear, SMART goals and how the standards relate to those goals, makes it easier to evaluate the standards themselves.

As the intent is for each chapter to be stand-alone, I'd recommend defining stream environment zones in this chapter.

This chapter refers to "adopted soil conservation targets." Why the change in terminology?

References are needed to support the statements that excessive impervious surface contributes to sediment and nutrient inputs to Lake Tahoe, and to support the statement that this leads to negative impacts on soil health, fisheries, etc.

Recommend describing, at least in general terms, what the land capability classes are based on. This could be done, for example, as a footnote to Table 5-1.

P 5-2: Nice to see improvement in data since 2011 allowing the determination of trend.

P 5-3: As I have mentioned elsewhere, please include the key to management/policy icons in Figure 5-1.

P 5-4: The definitions of hard and soft cover appear somewhat too rigid and not inclusive. What about surfaces such as gravel, which is covered (i.e. hard) but allows infiltration (i.e. soft)? Are these types of surfaces not extensive enough to warrant inclusion in a class?

P 5-5: Please clarify that LiDAR data were used to measure "actual" impervious cover in 2011. This was not immediately clear to me when I read it. It is also important to describe in this chapter with what frequency TRPA plans to re-use LiDAR to obtain updated impervious cover data.

P 5-8: Presenting the results for all land capability classes together was very useful to show where standards are attained and where they are not. TRPA might consider similar visuals for other chapters where there are a multitude of clearly similar and related standards (of course, this does not allow you to show trends, but could be complemented with trends graphs).

P 5-9: There is a reference to "unpermitted impervious cover." What does that entail? Can you provide any anecdotal or other information on how important this unpermitted cover might be in the Basin?

P 5-10: Good to see that the category with the worst status (Class 1b) is the one showing improvement.

I understand that there is high confidence in the accounting of acres of cover added. Are there no instances of decreases in cover? How confident are you that those data are accurate?

P 5-11: How effective is the physical removal of impervious cover? I realize that the removal is verifiable and contributes to the standard attainment, but would recommend discussion of how effective that removal is at allowing the site to improve the services it provides.

At the rate set by the interim target for class 1b (10 acres over 4 years) means that it would take 264 years to attain the standard. This suggests that something needs to change. Either the feasibility is too low, so the standard should be re-evaluated, or programs need to be developed and implemented to significantly increase the rate of improvement. What *is* feasible should be discussed.

The interim target for class 2 is unclear.

From the recommendations for the monitoring approach, it appears that you are considering an approach with two methods (LiDAR every few years, and permit accounting in the interim, calibrated to the LiDAR measures when those are taken), and regular ways of calibrating them to each other. Is this correct? If so, I support it, and would recommend you describe this more explicitly in this section.

P 5-13: The report clearly articulates the challenges with evaluating the Stream Environment Zones standard, which is very helpful in interpreting the rest of the chapter. I recommend that the description of the challenges be complemented up front with either specific recommendations of how to achieve them, or the presentation of the alternative ways of addressing these challenges, to provide decision-makers with a clear understanding not only of the challenge and potential solutions, but also the implications of choosing one approach over another. I realize that much of this has been done in the report. However, structuring it in the indicator summary makes it hard to fully understand the implications.

P 5-16: The fact that the SEZ restoration targets are percentage based, and the baseline used in this report to translate those percentages to "acres of SEZs restored" should be described in the Monitoring and Analysis section, not left to the Indicator State section.

The report states "the percentage targets are subject to change as the estimated extent of SEZ in the basin is revised based on new information." This is true even beyond the adoption of a "uniformly accepted SEZ map." Changes in hydrology driven by changing snowpack dynamics as temperatures warm, changes in water use patters as population grows, and other factors will likely continue to lead to changes in SEZs that may need to be captured in the baseline rather than in the current status of "restored acres." A significant investment in mapping and classifying SEZs in the Region may be needed to develop the "accepted baseline" for these standards. TRPA should carefully consider how to develop that baseline in a way that recognizes, accepts, and addresses the changes expected in the baseline as climate and other factors change. If the current opportunity to do so is missed, I would expect TRPA to face the same issue of a changing baseline in future reports.

P 5-17: How can new coverage in SEZ lands be "fully mitigated" when the standard is to preserve all existing naturally functioning SEZ lands? Or does the standard not imply that ALL naturally functioning SEZ lands should be preserved?

P 5-20: I fully support—and would like to highlight the importance—of the recommendation to develop a monitoring plan that enables assessment of SEZ condition. Given the use of EIP project tracking as the approach to monitoring the status and trend in SEZs, it will be important that the monitoring plan also determine what data EIP projects will need to report to allow TRPA to use these data to assess changes in condition over time.

Overall: As mentioned for other chapters, this chapter would benefit from a Conclusions section at the end, that summarizes the status and trends for the whole threshold category: what do the status and trends information on all these indicators say about the status and trends of Soils as a whole? And what are the overall recommendations for the Soils category?

Minor comments

P 5-15: Why is the map format different to the others in the report? Consistency helps the reader trace the linkages between different standards and chapters.

P 5-19: Why is overall confidence Low, when both confidence in both status and trend are moderate?

P 5-20: By when is the Taylor-Tallac area expected to restore 250-300 acres?

Chapter 7 – Fisheries

P 7-1: Very valuable to have TRPA's goal for the particular resources (fisheries in this case), as it provides much needed context for the thresholds selected. Though not specifically identified by Drs. Beauchamp and Naiman, their concern about the habitat-based standards not addressing the condition of the fish populations should lead TRPA to reconsider this goal: if the habitats being restored and monitored are not the main factors limiting fish populations, identifying those main limiting factors and revising the goal to target restoration and monitoring to improve those factors should be a priority for TRPA.

To help the reader track the relationships between indicators, it would be helpful to have a figure analogous to Figure 1-2, that details the full "tree" of categories, standards and indicators for the Fisheries category. This could be accompanied with a repeat Figure late in the chapter (see Overall comment at the end of this review) that also includes the reporting icon for each indicator.

P 7-4: Add to Figure 7-1 a simplified version of Table 1-1, the management and policy reporting icons.

P 7-6 (and all other maps in indicator summaries): Maps could be used more effectively to, for example, show the spatial relation between the location of fish habitat and the distribution of factors identified as Human and Environmental Drivers (e.g. urbanization along the shore zone); or is stream habitat in excellent condition adjacent to prime lake habitat? Additionally, *all* maps should be shown large enough to make information presented clear, and at a consistent scale and extent across all maps of the Basin, to allow the reader to further relate locations from one map (e.g. Figure 1-1) to locations in another map (e.g. P 7-6).

P 7-7: Is there data to estimate the potential impact of lake level fluctuations on the availability of "prime" fish habitat, especially spawning habitat? More broadly, I'd recommend TRPA make an effort to summarize and share with leadership available and up-to-date information on the human and environmental drivers identified, that should inform decisions and recommendations for changes. For example, if data exist on lake level fluctuations, and the frequency with which low lake levels occur during spawning periods, that would provide useful contextual information for interpreting the "somewhat better than target" status of prime fish habitat.

P 7-15: Why are site scores averaged in one monitoring period first, and then the trend assessed based on how these averages change over time? Isn't it possible that the condition of some of the sites is improving while at others it is declining? Such differences would be averaged out with this method. If there is a robust rationale for averaging site scores in one monitoring period first, please explain it clearly. If that rationale is not clear, consider assessing the trend over time at each site first, and then summarizing that information across sites. Averaging may not be the most informative way to do that, as the range and distribution of sites across that range is very useful information.

Related to the comment on P 7-7 on drivers: Another factor that would seem highly relevant in this context is occurrence of drought (identified as a driver on P 7-13 and P 7-15). On P 7-15 you state that low water levels are likely the largest contributor to poor condition in 4 of the 24 marginal sites, at least. Though it is true that if sampling occurred during wet years that proportion would decrease (as stated), it is also true that if droughts become more frequent in the future, then monitoring during wet years would be meaningless for determining condition for fish. I recommend including a summary of expected changes in drought frequency or intensity under a changing climate, to inform how this monitoring should be targeted.

P 7-16: Reference is made to the importance of projects to reduce stream temperatures. How is this dealt with? Recommendations for attaining or maintaining the threshold do not discuss this at all. Is it in the Recommendations chapter itself (not available at this time)?

P 7-19: The interim target attainment date paragraph concludes with two scenarios: persisting or worsening drought, or return of wetter conditions. Is there no information available on how likely these two scenarios are, especially as the climate changes? Understanding this is critical to interpreting this attainment date.

Stream Habitat Category: There is no reference or discussion of the effects of severe fires on sedimentation, riparian condition, etc. The Introduction chapter briefly discusses the legacy of fire suppression on forests in the Lake Tahoe Basin. Evidence is growing of the impacts of climate change on fire occurrence across western US forests. This context should be discussed in this report, to inform interpretation of stream habitat indicators, as well as considered for future monitoring, analysis and/or discussion.

P 7-20/23: What about the implications of climate change to stream flow? Given the importance of precipitation on stream flow, and the impact of increasing temperatures on snowmelt and the relative amount of snow and rain, what's known on climate change impacts on stream flows should be summarized, and the implications for future flows should be discussed.

P 7-22: About half way down the page: should "option" be "optimal"?

Overall: This chapter would benefit from a Conclusions section at the end, that summarizes the status and trends for the whole threshold category: what do the status and trends information on all these indicators say

about the status and trends of Fisheries as a whole? And what are the overall recommendations for the Fisheries category?

Missing reference: Byron et al. 1989

Chapter 8 – Wildlife

P 8-1: What is the wildlife goal TRPA is trying to achieve? Are they the two articulated in the Regional Plan (also on **P 8-3)**? Similar to some of the issues raised in the review of the Fisheries chapter, if the actions that are implemented affect habitat, but the goal is to sustain wildlife, there may be a disconnect if habitat is not the limiting factor.

P 8-3: Though it may be redundant with comments related to other chapters, I think the summary table of status and trends is very useful.

P 8-4: As I have mentioned elsewhere, please include the key to management/policy icons in Figure 8-1.

P 8-5: The discussion on deer was somewhat confusing. Why is this species called out for special treatment, though there is not an adopted standard for them? I expect there are other species that particular groups also track in the Basin. What makes deer different, and worth including in the report?

P 8-6, and more generally: Is the Tahoe Basin a large enough scale to be relevant for a population of large raptors? My concern arises because monitoring status and trends of a population at a finer scale than is relevant for the population as a whole may overlook or misunderstand the main factors affecting the population, which may be occurring—or occurring differently—outside the Basin. Partnering with entities that have the ability and the interest to monitor population status and trends at relevant scales would allow TRPA to evaluate if the Tahoe Basin components of each population are changing in similar way to changes in surrounding areas, and therefore be able to better inform actions taken inside the Basin to benefit these species. Please note that I make this comment as a generalist, not as a wildlife or avian monitoring specialist. Understanding the broader context may be sufficient (as noted in my next comment, for example).

P 8-7: Related to the point on relevant scale, the introduction to each standard (raptors) should describe the ecological relevance of the adopted standard (e.g. 12 population sites for goshawk), in the context of a goshawk population. For example, is there a maximum density (a carrying capacity) for each species for the basin? Are there habitat suitability models that could help estimate this?

The description of the impact of the Forest Service budget and monitoring emphasis on the availability of sufficient data for TRPA to evaluate status and trends of raptors in the Basin suggests that there is an opportunity for TRPA to strategically invest in monitoring that complements those of partners so that TRPA can, efficiently, meet its monitoring needs. I realize this may already be the case, and either funding or other limitations impede additional monitoring on TRPA's part.

P 8-8: The potential impact of fuel reduction activities on nests is identified under Effectiveness of Programs and Actions (also on P 8-11 for osprey). It is important to understand the trade-offs that exist between this risk, and the risk of habitat loss—even if temporary—due to wildfires. The goals for these raptor species, the standards to track progress towards them, and the expected benefits for the raptors need to incorporate such understanding. For example, should the number of population sites that need to be protected be larger, to allow for impacts of fuel reduction and wildfire? Or is there not enough habitat in the Basin to support further sites?

P 8-12: The analytical and monitoring approach recommendations appear to be contradictory. If a multi-year average is necessary because of the high year-to-year variability, then changing the annual method due to lower need to invest in monitoring must be carefully considered. Or is the proposed monitoring approach to monitor

multiple consecutive years (allowing for a multi-year average), then not monitor for another sequence of years, and so on?

It is important to ensure that the revisions to the threshold standard to clarify the standard's objective not only reflect best available science, but must also clearly respond to TRPA's goal/s for wildlife in the Basin (which should also be clarified).

P 8-17: As stated for other chapters, applying the same criteria for trend confidence to a trend of "little or no change" as are applied to "improvement" or "decline" trends is counter-intuitive. If the regression analysis suggests there is no trend, then of course the r2 will be low and the p-value will be high. These metrics reinforce the lack of trend, rather than undermining the confidence in the result that there is no change. I recommend consulting with a statistician to evaluate how the r2 and p-values should be interpreted to inform the confidence measures.

P 8-18: You make a good point about the need for consistency in definitions across indicators and across species. Articulating the chain of logic that links species and indicators to a greater goal for wildlife in the Tahoe Basin can help set a foundation for that consistency.

P 8-19: Though I do not argue against the importance of 75 deaths per year in wind farms, this value on its own does not provide sufficient context as to what that impact means. How large is the golden eagle population in the state of California? How does that relate to this mortality value? Are wind farms an important threat to golden eagles that nest in the Tahoe Basin?

P 8-26: Under Relevance, you articulate waterfowl's value as "an indicator of the biological integrity or relative value of a site for providing a functioning wetland ecosystem." This is an example of why it would be important to relate indicators in one category to other categories, and to design a comprehensive program that minimizes redundancies and inefficiencies. If overarching goals tie different standards together, then it will be easier to identify where some standards—such as waterfowl—might function effectively as surrogates for other standards—such as wetland areas. These links need to be described in the report.

P 8-29: I fully support the Recommendations under Monitoring Approach.

P 8-32: Are the programs and actions effectively maintaining the "disturbance-free" condition of these designated zones? I would recommend that TRPA consider shifting to (or adding) a numerical target to this indicator, to capture whether these zones—which are mapped—are currently disturbance-free.

P 8-34: As mentioned in my review of other chapters, there is important contextual information and data on the human and environmental drivers that TRPA should provide with the standards, to better understand the implications of the status and trends. In this case, what is known about climate change impacts on wetlands? Does that suggest changes are needed in actions, in the standards themselves, or in monitoring?

The importance of the condition of riparian and wetland areas is clearly identified in this chapter, yet there is no standards to track whether these protected riparian and wetland areas are in sufficiently good condition to provide the wildlife and other values that is expected of them. Again, explicitly linking between categories and indicators might allow the report to better inform decision-makers on such condition (e.g. with waterfowl, or soils, or water quality).

P 8-41: The recommendation to explore coordination with land management agencies outside the basin to improve habitat connectivity and migration routes is commendable, and the need to do so should be considered for all wildlife species that might have important out-of-basin drivers.

<u>Minor comments</u>

P 8-8: Please note where the references are personal communications. E.g. Zanetti 2015.

P 8-14: Please add to the map on P 8-13 those locations where biologists set up observation points. This would help understand the design, and get some sense of whether that design is sufficient.

P 8-25: Please accompany the average rating of human disturbance with some measure of variation across sites.

Chapter 9 – Scenic Resources

This chapter clearly articulated how the scenic resource values across the Lake Tahoe Basin were defined, and that they had each been identified and rated for quality in 1982. This comprehensive definition of what values are being maintained and enhanced provided a solid foundation for the most recent rating of those same values, and the evaluation of trends. Therefore, my comments on this chapter are generally minor.

Minor comments

The three appendices should be referenced in the chapter. In addition, the order the information is provided in Appendix 2 should be consistent with the order it is given in the chapter – roadway travel units first, followed by shoreline travel units.

Given that the ratings of the different scenic resources indicators are subjective, a brief description of how and why the approach is credible and repeatable would be valuable.

P 9-1: How do you define "the *vast majority* of views" referenced in the first goal?

P 9-2: The first row in Table 9-1 refers to Tables in the Draft Study Report, but the full reference of this Draft Study Report is not available. Similarly, the second row refers to the 1993 Lake Tahoe Basin Scenic Resource Evaluation. I recommend using a standard format for references, as you have done in other chapters, and include a References section at the end of the chapter.

P 9-4 (and pages summarizing status and trend determinations for the other standards): I felt it would be clearer for the reader to organize this summary with the different status categories across different columns, and the different trends categories in different rows. This would visually convey how many units were in and out of attainment, and whether there is a relationship between those out of attainment and the direction and magnitude of change. For later tables, each cell in the body of the table could have the icon plus the number of units that fell in that combination of status and trends. Such an approach could also be applied to other chapters, such as the Soil Conservation chapter, that has a set of related standards (impervious cover in different land capability classes) analogous to those for the different travel units in this chapter.

P 9-7: I was left wondering why a 0.17% per year increase was considered a "marked improvement", particularly since later in the chapter (**P 9-17**) a 0.19% per year decrease was described as "statistically insignificant." The language used in a number of the trend summaries implies that the authors consider the changes to be more important than the standard methodology of this report reflects. If that is indeed the case, I would recommend you state that, and explain why.

P 9-8 (also repeated under the other standards): I fully support the approach described under Attain or Maintain Threshold, of using these data to update the Scenic Quality Improvement Program, and encourage the TRPA to prioritize their efforts and investments guided by these data.

P 9-11: The messages conveyed by the pie chart and the map are visually contradictory. I realize that the travel units are not of standard lengths, and there is a good reasoning for that. However, this has implications in terms of the absolute progress needed to reach full attainment, especially if the 31% of units out of attainment are significantly longer than those in attainment. Given that how the travel units are defined is clear and consistently applied, I do not recommend changing the indicators or approaches in any way, rather I

recommend discussing the implications of the different unit lengths, especially in light of the low rates of improvement observed so far.

P 9-13: In line 4 you refer to a table. Should this be "the graph below"?

P 9-15: Provide a couple of key references included in "characteristics are well documented in the academic and professional literature…"

Give a brief definition of the characteristics measured – unity, vividness, variety, intactness. Level of detail should be analogous to that provided for coherence, condition, compatibility and design quality on **P 9-25**.

It would be helpful to the reader if you provide some examples of what are the "scenic resources" identified in the 1982 inventory.

P 9-16: Why do you provide regression lines and equations for some of the Trend graphs, and not others? Recommend consistency, and given that you discuss the trends in each indicator summary, I support including the regression information in each.

P 9-23: Why is there a single regression shown, rather than a split regression (before and after 2001), as is discussed in the last paragraph on **P 9-22**?

P 9-25: What are class I and class II bicycle trails? Briefly describe.

This summary explicitly describes how the baseline condition was established for scenic resources added later than 1982. I did not see that described for the earlier indicators, so would recommend including that.

Why is the Other Areas indicator reporting category only compared to the baseline, rather than also having a minimum score? Is it because baseline in all cases was well above 50% of the maximum score? I would recommend describing why it is different (unless there isn't a robust rationale).

P 9-31: There is no reference to number of buildings. I realize that this policy is focused on design elements of buildings that are being built, but the accumulation of buildings will also affect scenic quality, and therefore may need to be considered. Is that sufficiently covered under other indicator categories, such as travel route ratings? If that is the case, it would be worth stating that.

P 9-34: Last line appears incomplete: "… modifications that could clearly community character."

Chapter 10 – Noise

P 10-1: Is the goal to "address noise impacts to both wildlife and visitors"? Articulating what TRPA strives to achieve by addressing those impacts would help determine if goals are achieved.

P 10-2: Some context on why so many single noise events standards are necessary would be useful. I realize that knowing the source of the noise event is useful in determining how to address the issue. However, in terms of whether noise events are a problem or not, so many standards seem excessive. In addition, isn't the frequency of exceedances important?

P 10-13: Why is the recommendation to wait until the new shorezone plan is implemented to re-engage monitoring? Having pre-implementation data would allow you to evaluate the effectiveness of the plan.

P 10-14: Please explain in the description how the monitoring locations were selected, and why.

Please also define "cumulative noise events." These are accumulated across sources of noise within the same period of time, rather than an accumulation of single noise events over time?

P 10-15: I suggest graphing the data with a year-by-year x-axis, to both visually show the spread of years with data, and also to ensure it is clear that the regression line has year as the independent variable. The current graphs left me uncertain how the regression was developed, as you would expect the line, which would be straight if all years from 1991 to 2015 were shown, to have breaks, as the timeframe shifts (e.g. 1991-2011 has the same x-distance as 2014-2015).

P 10-17: In the interest of clarity and transparency on the review process, I would recommend using a different term (not "outside peer review for the 2016 Noise Threshold Evalutions") for Ascent Environmental's feedback on the Noise chapter. This was not part of the peer review process we are currently carrying out, and Ascent Environmental is one of the contributing organizations to the Noise chapter (see Exclusion List). Clearly Kerr et al.'s input is significant, and TRPA should be commended for ensuring they receive this input in developing the Noise chapter of the Threshold Evaluation. It is just different to the "outside peer review."

P 10-18: What is the difference between the map for Wilderness and Roadless Areas, and Critical Wildlife Habitat? If they are not different, does it make sense to have two thresholds?

P 10-19: Why are site scores averaged across sites in one monitoring period first, and then the trend assessed based on how these averages change over time? Isn't it possible that noise levels at some of the sites are improving while at others they are declining? Such differences would be averaged out with this method. If there is a robust rationale for averaging site scores in one monitoring period first, please explain it clearly.

P 10-29: Interim Target and Target Attainment Date state threshold is in attainment when it is not (also P 10-47, maybe others).

P 10-43: Why were there no measurements in 2015? Explain under Status.

General comment for Cumulative Noise Events: Isn't the location where CNEL was monitored for each land use type before the monitoring protocol changed still part of the current monitoring? If it is, you could look t trends using the same data as measured before the protocol changed, and then compare that one-site value to the more robust results using the new protocol, to determine the difference in values caused by the change in protocol.

General comment for Cumulative Noise Events for the different transportation corridors: Would the programs and actions to address transportation noise be different in different parts of the Basin? If the approach to addressing issues is not corridor-specific, TRPA should evaluate ways of streamlining the monitoring.

P 10-60: On graph, show when the monitoring protocol changed, as you have in other graphs in this chapter.

P 10-62: Was the threshold in attainment in 2015? Don't you use the maximum 24-hour CNEL dBA, shown in the graph with the error bars?

Chapter 11 – Recreation

It was interesting that this is the only category that has only policy statements as standards. It was clear from the chapter and indicator summaries the challenges posed by the lack of numerical standards. This both reinforces the need for the TRPA to consider how to move from policy and management statements to numerical standards, and highlights the particular importance for doing this around Recreation.

P 11-1: Very valuable to have TRPA's goal articulated. The language used ("to achieve the recreation threshold standard") presents the standards as the end itself, rather than as a means of tracking progress towards the ultimate end of "ensur[ing] equilibrium between the region's natural endowment and its manmade environment."

Briefly describe where the "best available information" was obtained.

P 11-2: The graph showing the results of the satisfaction survey suggests that a framework already exists among monitoring partners to set numerical targets for recreation. Why has it not been done?

P 11-3: Why isn't one of the evaluation criteria a question such as in the Scenic Resources chapter: "Is there evidence to suggest these actions are achieving the intent of the policy statement?" I realize that the evaluation of surveys and review of land acquisitions and development of public access amenities may be addressing that, but it is valuable to make that explicit.

P 11-4: Somewhere in this indicator summary there needs to be a clear discussion of the uncertainty and limitations to the conclusions drawn from the surveys because they are incomplete, opportunistic, and have such different objectives that their aggregation is questionable. The way the information is organized, there is recognition that the data is not the best, but the results are then presented as "truth", without discussion of the impact of data quality for this purpose, or their ability to effectively function as "evidence that actions are achieving the intent of the policy statement."

Why aren't values provided on percent of total land area held in public ownership under Status?

Overall, having policy statements with no numerical targets means that status doesn't provide much information about what is happening with the resource of interest, rather it informs whether action is being taken. Trend in the values reported here would be much more valuable in informing whether those actions are leading to positive change, or whether the actions are not being effective. Therefore, assessing trend should be a priority for quantifying evidence of effectiveness of policy statements.

P 11-6: If data such as those presented in Figure 11.1 were available for 1982 or other earlier date, it would be possible to establish a baseline for these resources against which to compare the current status (analogous to the baseline developed for scenic quality resources – subjective, but standardized and consistent across space and time). Is this possible? If not, I recommend evaluating whether the 2012 values could provide such a baseline, against which to compare these and future status.

P 11-7: I was encouraged to see the recommendations to the monitoring approach, and fully support all three, especially the third: coordinating and combining the recreation monitoring with the scenic quality monitoring.

In contrast to my last comment, I was surprised that no changes are recommended for the standard itself. As you have mentioned in other chapters, and as I highlighted at the beginning of this chapter's review, not having numerical targets makes monitoring challenging. And in this case, you have at least some data suggesting what numerical targets would allow you to assess effectiveness of actions taken in response to this policy statement.

It is encouraging to see the recognition that climate change may have impacts that affect the amount, quality, and distribution of recreational opportunities, as well as the patterns of recreational use. I encourage the TRPA to further evaluate existing information on climate change and climate change impacts, and further develop and plan for such changes.

P 11-9: The distinction between improvement and creation of recreation facilities should be discussed. Given that the majority of progress is in improvement rather than creation, how much impact does that have on recreational opportunities and/or the quality of the recreational experience?

P 11-10: The TRPA should consider a two-method monitoring approach, analogous to what is started (and proposed for the future) in the Soil Conservation chapter: combine the annual tracking of permitted projects with an independent measure of what is occurring across the Basin. This may not be as straightforward as the use of LiDAR to measure actual impervious cover across the Basin, but is nonetheless important to include.

P 11-11: The PAOT system was unclear. Granted this is not a system I have experience with, though I would argue that the same might be the case for interested stakeholders looking to better understand the progress the TRPA is making.

Peer Review of the Tahoe Regional Planning Agency's 2015 Threshold Evaluation Report

Individual Reviews of "Recommendations for Review and Modification of Thresholds Standards"

September 23, 2016

Submitted to:

The Tahoe Regional Planning Agency

By:

Sonia A. Hall, PhD – Associate Scientist

Christine M. Albano, MSc – Lead Scientist

and

Brett G. Dickson, PhD – Chief Scientist

11050 PIONEER TRAIL • SUITE 202 • TRUCKEE, CALIFORNIA • 96161 • (p) 530.214.8905 | (f) 530.214.8907
5 OLD TOWN SQUARE • SUITE 205 • FORT COLLINS, COLORADO • 80524 • (p) 970.484.2898 | (f) 970.484.1069
info@csp-inc.org | www.csp-inc.org

Table of Contents

Note: In addition to their individual review, one reviewer also provided detailed comments in an annotated version of the chapter, which was submitted to the Tahoe Regional Planning Agency. For simplicity and consistency, however, the annotated chapter was not included in this compilation of individual reviews.

Biosketches – Peer Review Expert Panel

Madeleine McKinnon, PhD – Conservation International (now Vulcan Philanthropy)

Dr. Madeleine McKinnon is currently Senior Director for Monitoring and Evaluation in the Moore Center for Science at Conservation International. She leads monitoring and reporting of CI's institutional achievements across its global programs. Madeleine provides scientific leadership to the organization on applications of performance measurement, impact evaluation, and evidence synthesis. Originally from London, she graduated from the University of Edinburgh and has a master's in applied ecology and conservation from the University of East Anglia. She earned a PhD in program evaluation from the University of Queensland, Australia. She has published on aspects of conservation evaluation, systematic conservation planning, and prioritization of conservation investments. She serves on the Board of the Conservation Measures Partnership and is a principal investigator of a Science for Nature and People partnership (SNAPP) working group on evidence-based conservation at the UCSB National Center for Ecological Analysis and Synthesis (NCEAS). She will shortly be leaving CI to join Vulcan Philanthropy as Director for Measurement and Impact. She will be responsible for monitoring and reporting on the outcomes and impact of Paul Allen's, cofounder of Microsoft, philanthropic investments.

Andy Rowe, PhD – ARCeconomics Ltd.

Andy Rowe has worked for over thirty years as an economist and evaluation consultant. His evaluation work focuses on sustainable development, natural resource management and dispute resolution settings. He has a PhD from the London School of Economics, is an associate editor of Evaluation and Program Planning, and a Fellow and former President of the Canadian Evaluation Society. He has published several articles and book chapters on evaluation and sustainable development, has developed new evaluation methods for evaluating sustainable development, and has developed a grant making strategy for use-inspired science research with Kai Lee of the Packard Foundation. His Rapid Impact Evaluation approach is being piloted by the Treasury Board of Canada for inclusion in the Canada National Evaluation Policy and applied by the Global Environment Facility. He is a member of the Board of Ecotrust Canada.

Nick Salafsky, PhD – Foundations of Success

Nick Salafsky is Co-Director of Foundations of Success, a non-profit organization that seeks to improve the practice of conservation. Foundations of Success has worked for over 15 years with conservation practitioners around the world to define clear and practical measures of conservation success, determine sound guiding principles for using conservation strategies, and develop the knowledge and skills of individuals and organizations to do good adaptive management. Nick is also product manager of the Miradi Adaptive Management Software program and Co-Chair of the Conservation Measures Partnership, a community of practice composed of many of the world's leading conservation organizations and agencies.

Prior to starting FOS, Nick worked for the MacArthur Foundation where he was responsible for environmental grantmaking in Asia and the Pacific. Nick also worked for the Biodiversity Support Program, testing enterprise-based approaches to biodiversity conservation across the Asia/Pacific Region. Nick spent several years in West Kalimantan, Indonesia, conducting interdisciplinary research on the forest gardens, a locally developed agroforestry system, and the behavorial ecology of the red-leaf monkey. Nick has a Ph.D in Environmental Studies and an MA in Resource Economics from Duke and an AB in Biological Anthropology from Harvard.

Review by Dr. Madeleine McKinnon

Conservation International; Vulcan Philanthropy

Thank you for inviting me to act as a reviewer of the Assessment chapter within an important and timely report. In this review, I provide general comments on the chapter below which correspond to the questions for peer review considerations. In addition, I have made detailed edits and comments on to the electronic PDF document in a series of comment boxes. I hope that these are easily interpreted.

General comments

The chapter proposes a critical first step in a substantial process to validate, prioritize and update existing threshold standards for monitoring and evaluation of the Lake Tahoe Region. Overall, the chapter describes a revised approach to assessing standards which mirrors similar ongoing efforts by other organizations, agencies and philanthropies in the environment and natural resources sectors. The proposed approach outlined in the chapter offers an objective and timely framework. It is generally in alignment with existing best practice. I commend the TRPA for embarking on this assessment which should allow more informed and defensible prioritization of standards and a more efficient and sustainable monitoring framework for future evaluations.

Overall, the chapter requires substantial revision for clarity on terms, rationale and process. These revisions are necessary to improve readability, ensure consensus on the process and to adequately support the proposed approach.

The use of terms is inconsistent – goals and objectives are used interchangeably, framework is used at different scales and for different purposes, and some terms (e.g., results hierarchy) are not adequately defined.

The headings and subheadings used are confusing and should be revised to improve the structure of the chapter. The introductory section does not clearly orient the reader to the purpose, scope and key topics described in the following chapter. (I found the peer review guidance note more informative). The chapter ends abruptly without a clear set of conclusions or initial next steps.

A suggested set of headings and subheadings is listed below which can be adopted or revised as authors feel appropriate:

1. Introduction
    a. Summary of threshold standards and evaluation report
        E.g., how do recommendations in overall report warrant a new assessment approach?
    b. Purpose of this chapter
    c. Overview of chapter sections
2. The need for assessment
    a. Origins of threshold system
    b. Issues and inadequacies of current system
3. Proposed methodology for assessment
    e.g., describes the two main parts of the assessment framework plus the process for implementing it (participatory approach?)
4. Caveats and challenges
5. Next steps/moving ahead

The two proposed steps in the assessment process (SMART criteria and categorization questions) are not consistently described nor adequately referenced from existing literature from which these approaches are borrowed. The description and rationale for SMART criteria is generally fine, but the same treatment should be

applied for the categorization questions. The logic and description for choosing and applying each step should be the same, e.g., definition, rationale, approach and example.

Specific comments (see also annotated PDF)

1. *Sufficient background on threshold standards*

The chapter does not stand alone and it was necessary for me to review the summary chapter in order to understand the key definitions and rationale for threshold standard system. The chapter does document the origins of the standards system, but this is less useful than an understanding of what the threshold standards are now (e.g., the existing types of standards – leisure, wildlife etc). The historical perspective is only appropriate in that it should provide an account of how standards were identified and selected in previous iterations and why previous approaches are not optimal now and thus should be revised (using the assessment framework described in the chapter). As it is currently written, the reader is immersed into a review of the system origins without context for why this information is relevant or contributes to the issues discussed in the rest of the chapter. Description of the origins should thus primarily be about providing context for the proposed assessment.

2. *Sufficient background on best practice on setting goals?*

The chapter does a fair effort in describing best practice around the proposed approach for better defining goal statement (i.e., application of SMART criteria). The authors cite several international organizations using SMART criteria to set goals. It might further strengthen this section to describe U.S. national or state processes that are more similar to the Lake Tahoe regional context than broader international processes such as the GEF which have a very different scope. An example from one of the US-based examples would also be useful.

3. *Clear rationale for the proposed assessment?*

The introduction to the chapter is poorly written and does not clearly orient the reader about the objective of the chapter and the topics to be covered to achieve this objective. The introduction to the peer review guidance letter is actually more informative than the opening page of the chapter. It becomes clear over the course of the chapter what the purpose of the assessment is, but it should be better stated up front. This is particularly relevant for a such a large report where readers are likely to be quickly glancing at individual chapters.

4. *Proposed assessment framework logically consistent and clearly articulated?*

The rationale and description of the SMART criteria is generally clear. The use of the example is helpful. I would recommend that the table (13.1.1) is embedded earlier in the chapter with the text corresponding to SMART criteria rather than at the end of the section.

The rationale for applying the "categorization questions" is less and selection of this framework is poorly linked to best practice or existing guidance. The concept or term "categorization" is not widely used and does not have broad currency in the relevant sectors. I understand what the authors are attempting to propose in terms of the assessment, but this framework should instead be framed as a prioritization exercise based upon "the value of information" and "the ease of generating that information" for monitoring and evaluation purposes. The authors might refer to several publications that examine this issue:

McDonald-Madden E, Baxter PW, Fuller RA, Martin TG, Game ET, Montambault J, Possingham HP. Monitoring does not always count. Trends in Ecology & Evolution. 2010 Oct 31;25(10):547-50.

Jones JP, Asner GP, Butchart SH, Karanth KU. The 'why','what'and 'how'of monitoring for conservation. Key Topics in Conservation Biology 2. 2013:327-43.

Maxwell SL, Rhodes JR, Runge MC, Possingham HP, Ng CF, McDonald-Madden E. How much is new information worth? Evaluating the financial benefit of resolving management uncertainty. Journal of Applied Ecology. 2015 Feb 1;52(1):12-20.

The criteria for categorization does not need to change, but the context for the categorization questions would be better situated with reference to "value of information" and "prioritization of resource allocation" concepts.

5. *Proposed approach for assessing the threshold standards consistent with best practice in the field of monitoring and evaluation? Is the proposed approach technically sound?*

The range of criteria utilized in the assessment is generally in alignment with best practice though does not reflect the most recent scientific thinking. The current approach still involves considerable subjectivity in how different standards will be prioritized. It is not clear how responses to criteria within the assessment framework will be further assessed, weighted and ranked in terms of importance of each standard. Are criteria equal? I assume a participatory approach will be implemented, involving relevant stakeholders, to assign values and prioritize across standards. A section briefly describing this participatory process would be useful.

Recent thinking suggests assigning quantitative values where possible to different criteria, e.g., cost of monitoring, and attempting to quantify the benefit function of different standards that reflect the value of the conservation feature. This approach was originally developed to prioritize species management actions and conservation, but has recently been developed and applied for prioritizing monitoring efforts more broadly.

6. *Are there additional questions that you think would help decision makers prioritize investments in updating the threshold standards?*

As suggested above, a revised approach which is more explicit about the value of information generated by different standards would be desirable. This is more consistent with best practice and current scientific thinking than the vague notion of "categorization".

I have included several additional questions below and a brief rationale for their consideration in the framework.

a. Is the standard established and/or broadly applied?

it would be useful to know whether the standard is established and utilized more broadly beyond the Tahoe region for alignment purposes and efficient data collection and synthesis. This would align the standards potentially with state, federal and international standards for monitoring and reporting, e.g., indicators used to report on US EPA Act, UN Convention on Biological Diversity and other agreements. Furthermore, standards with broader adoption might already have existing secondary datasets which could be utilized and thus be more efficient (less costly).

b. Is the standard based in current theory?

In part this question is already couched in the "science-based" category, but I think it could be useful to disaggregate "theory-based" and "evidence-based". The first is whether the standard is plausible based on our current understanding of theory (e.g., the species-area relationship assumes the larger an area, the more species occur) and the evidence exists to empirically demonstrate our assumptions (e.g., results from scientific studies estimate the number of species likely to occur in different sized areas).

The selection of standards should thus correspond to a hypothesis of a projected outcome and the effects likely to be observed. Current best practice recommends the development of a conceptual model, or theory of change, which illustrates the activities, intermediate and long term outcomes for a conservation strategy. Indicators are measurable statements used to track components of a theory of change. Ideally, each standard or groups of standards would be mapped on to a theory of change – e.g., a conceptual model that illustrates how a set of management and policy activities will achieve the desired results. The CMP Open Standards, already

referenced in the document, provide detailed guidance on how to construct such conceptual models, or results chains as CMP refers to them.

7. *Budgetary constraints mean that it may not be possible to review the entire suite of 171 standards all at once. Will the proposed assessment provide information necessary to prioritize review efforts? Are there additional pieces of information that may help inform the prioritization of scare resources?*

See comments above about assigning more explicit cost values to monitoring efforts. There are several approaches and tools to assist prioritization including multi-criteria decision analysis in prioritization of monitoring can be determined using a fixed budget. It is generally not preferable to use an arbitrary scoring approaches as this might over- or underestimate the relative importance of different criteria.

For more information, see articles below.

Klein CJ, Jupiter SD, Possingham HP. Setting conservation priorities in Fiji: Decision science versus additive scoring systems. Marine Policy. 2014 Sep 30;48:204-5.

Field SA, Tyre AJ, Possingham HP. Optimizing allocation of monitoring effort under economic and observational constraints. Journal of Wildlife Management. 2005;69(2):473-82.

8. *General impressions of the TRPA standards: This is your opportunity to provide general feedback on standards. Are they similar to other systems you have worked with? Are there points of difference from other systems you have worked with? If you were charged with leading the review and update of the standards, how and where would you start the process? Is there any guidance you would like to provide the agency and its partners as they begin the review process?*

The TRPA standards are not dissimilar to other systems that I have worked with. They do reflect the tendency of many monitoring systems to involve a long “laundry list” of indicators without careful consideration about the value of information and how this information is intended to inform decision making.

The structuring of standards as management, policy etc is a different organizational framework than other systems I have worked with which have tended to be organized using a “state, pressure or response” framework or a hierarchy such as “inputs, process, outputs, outcomes” reflecting stages of the project management cycle.

The new categorization questions will hopefully provide a more useful typology for organizing the standards in a more intuitive manner.

In my experience, it is useful to have consistency in the general domains (and subdomains where relevant) of standards and ensure agreement on these, but allow some flexibility on the specific indicators in particular if these might be measured at different scales for different purposes by activities across the region at different time points.

Review by Dr. Andy Rowe

ARCeconomics Ltd.

Introduction

To prepare this review I have considered the direction provided in the *peer review charge* and the contents of *Chapter 13: Conclusions and Recommendations* and of *Threshold Standard Assessment Briefing material*. My review addresses the questions identified in the peer review charge:

- Is sufficient background provided in the chapter such that a reader not previously familiar with the threshold standards can understand the history of the threshold standard system?
- Is sufficient background provided in the chapter such that a reader not previously familiar with best practice for setting goals can understand the potential utility of comparing the existing system with best practice?
- Does the chapter clearly lay out the rationale for the proposed assessment?
- Is the proposed assessment framework logically consistent and clearly articulated?
- Is the proposed approach for assessing the threshold standards consistent with best practice in the field of monitoring and evaluation? Is the proposed approach technically sound?
- Are there additional questions that you think would help decision makers prioritize investments in updating the threshold standards?
- Budgetary constraints mean that it may not be possible to review the entire suite of 171 standards all at once. Will the proposed assessment provide information necessary to prioritize review efforts? Are there additional pieces of information that may help inform the prioritization of scare resources?
- General impressions of the TRPA standards: This is your opportunity to provide general feedback on standards. Are they similar to other systems you have worked with? Are there points of difference from other systems you have worked with? If you were charged with leading the review and update of the standards, how and where would you start the process? Is there any guidance you would like to provide the agency and its partners as they begin the review process?

While I was not requested to provide ratings on each of these questions I have done so on most of the questions to ensure that my assessment is clear. I have used the categories of highly satisfactory, moderately satisfactory, satisfactory, moderately unsatisfactory and highly unsatisfactory.

Background Material

Chapter 13 has extensive background material including a brief history and definition of the threshold standards and the reason for the current modifications. This provided me with an adequate first introduction to the threshold standard system including the rationale for the system, political support, the shift in the approach over the years and the reasons for the current modifications.

I regard the background material as fully satisfying the needs of a reader with some familiarity with environmental quality approaches.

Best Practice in Setting Goals

I have been unable to find a rationale for using best practices in the assessment or for issues such as the criteria for selecting the best practices to review. I assume this is covered in other chapters but a sentence or two on

the rationale for selection would be useful. For example I found myself wondering how many of the twenty best practices reviewed are germane to standards for Lake Tahoe and how the best practices inform the review. A concluding chapter is not the place for an extensive discussion but it should include sufficient description to provide some assurances that this was a systematic and relevant review. Readers who were not familiar with best practices in environmental management are not provided with any explanation of what best practices are or why and how they are used. The Lincoln quote was a distraction.

I regard the chapter as moderately unsatisfactory in providing the less well informed readers with an adequate understanding of the potential utility of comparing the existing system with best practice.

Rationale for Proposed Assessment

The Chapter provides a succinct description of the authorities for the review. It would be helpful for readers to be able to connect the numeric list at the top of page 13-4 with the rationale for this review (is it just the five year duration or were one of the criteria breached?).

The first point in the numbered list in the next paragraph on page 13-4 states: *1. Developing rationale and a proposed action for the amendment.* The subsequent paragraph provides some rationale for the approach to the assessment but it is more a description of the approach than an articulation of the rationale for the approach.

The rationale for the review itself and for the approach taken by the review do not provide a very useful reference point should one want to reflect on whether the review has accomplished the intended goals. Coverage by the chapter of the rationale is in essence to say the authority exists and this is a quick overview of how it is being addressed. Rationale should also address the question of *why,* inform expectations of what the review will provide and enable backward reflection and assessment of whether the assessment has fulfilled its purpose and also provide essential reference points for adaptive management of the standards.

I regard the description of the potential utility of the rationale for the assessment and of the assessment approach as moderately unsatisfactory.

Consistent and clear articulation of the proposed assessment framework

The material in Chapter 13 consistently and clearly articulates the proposed framework. Use of examples assists understanding. Some examples of material that could improve this include:

- One or more figures would be very useful (e.g. for bottom 13-6 top 13-7)
- Clarifying how temporal and spatial scales will be included (perhaps top 13-6)
- A copy edit (e.g. phrases such as *type differentiation is included* can be difficult for some readers to understand 13-6)
- How the assessment addresses relative importance and connectivity of indictors (it is described in the *Threshold Standard Assessment Briefing material* but not in this chapter)

I regard the chapter as providing a satisfactory (consistent and clear) articulation of the proposed assessment framework. It will benefit from a good copy edit including appropriate attention to the reading level of this section of the chapter.

Alignment with best practices in monitoring and evaluation

The chapter does not address monitoring and evaluation let alone best practices. The major reference to the M&E field is Stufflebeam and Shinkfield (2007) a textbook that provides an assessment and description of a number of different evaluation approaches. I suspect that all of the material in this volume will refer to human systems. There is material on evaluation in natural systems including coupled human and natural systems. Stufflebeam and Shrinkfield is a throwaway citation that does not support any of the questions to which it is applied on 13-8 and 13-9 and is not relevant to this undertaking.

To approach adequacy on M&E the chapter would be best to have clearly described the rationale for the standards (see above), the difference that is sought through revision of the standards and that are expected from the shift in the approach to standards. These should be described as outcomes not just outputs and should be as observable as possible. The chapter should also refer to how implementation of the standards will be monitored including timing and approach and how the monitoring information could support adaptive management of the standards and be used to assess success against the rationale and the expected contributions of the standards to management decisions and other purposes.

I regard the extent and manner in which the chapter addressed monitoring and evaluation as highly unsatisfactory.

Additional questions to help decision makers prioritize updating the threshold standards

The literature on use of global environmental assessments identifies features of assessments that improve prospects that they will be used and be influential[1]. We now know that the assessment process and in particular engagement of key interests in the assessment positively influences use. We also know that timing and salience are formative issues. The proposed assessment would benefit from considering how key interests (e.g. decision makers, user and affected communities, tribes) could engage with the assessment including definition of (additional) questions that the assessment should address and priorities. The assessment should also consider the timing of decision opportunities for using knowledge from the assessment.

Prioritization of scarce resources for the standards

This is not addressed in Chapter 13. The literature on use of assessments points to a prioritization that is not entirely driven by technical priorities and which include the priorities and questions of decision makers and key stakeholder interests.

General impressions of the TRPA standards

I very much like the particular scorecard approach described in the briefing materials. These capture the important elements and are readily used. I am concerned that processes for adaptively managing the standards have not been articulated and that with budget and other pressures that it is likely that they will not be adaptively managed.

Overall assessment

The current chapter has shortcomings, some of which I regard as serious and others that can easily be addressed. My main concern is that the overall approach does not incorporate good practice on undertaking assessments in a manner that promotes use.

[1] For a recent summary see Clark, William C., Lorrae van Kerkhoff, Louis Lebel, and Gilberto C. Gallopin. 2016. "Crafting Usable Knowledge for Sustainable Development." *Proceedings of the National Academy of Sciences* 113 (17): 4570–78. doi:10.1073/pnas.1601266113. Also Rowe and Lee (2012), *Linking Knowledge with Action*, Packard Foundation https://www.packard.org/wp-content/uploads/2013/05/LinkingKnowledgewithAction_ScienceCS2013.pdf

Review by Dr. Nick Salafsky

Foundations of Success

Responses to Questions in Reviewer Charge

Overarching Note to this Review

For better or worse, I am heavily influenced by the Conservation Measures Partnership's *Open Standards for the Practice of Conservation*. From my (admittedly biased) perspective, one advantage of the *Open Standards* is that they provide a very precise and clearly defined set of terms for describing the components of this kind of performance management system (e.g. the difference between a focal target, a goal, an indicator, a measurement, and a desired future value).

Although in this review I have tried to adopt the terms that the report itself uses, in places I have had to fall back on using the *Open Standards* lexicon as a reference point to try to clarify subtle distinctions in the TRPA Threshold Standard system. TRPA does not necessarily need to adopt the *Open Standards* terms per se – but it does have to have one set of terms that are well defined and clearly understood by all involved parties.

#1. Is sufficient background provided in the chapter such that a reader not previously familiar with the threshold standards can understand the history of the threshold standard system?

Overall, the section on the introduction / history of the standard system is both clear and succinct.

#2. Is sufficient background provided in the chapter such that a reader not previously familiar with best practice for setting goals can understand the potential utility of comparing the existing system with best practice?

Per my note on terminology above, there is substantial confusion throughout this chapter about the differences between:

a. The overall threshold category or threshold standard being assessed (e.g. water quality in Lake Tahoe or water clarity in Lake Tahoe),

b. The more specific management goal that has been set (e.g. reduce suspended sediment levels by 15% from 1990 levels)

c. The standard indicator/methodology being used to measure the goal (e.g. Secchi disk readings at specific locations at certain times each year)

d. The desired future value of the measurement (aka target level) that relate measurements to the management goal (average Secchi disk visibility at no less than 29.7 meters). In some cases, these values can be baked directly into the goal statement in Point b, but in other cases they may be a more specific articulation of the goal.

More specifically, in Chapter 13 when we talk about assessing *threshold standards* it is not clear how this concept relates to the rest of the report as shown in Figure 1-2 below. Is the *threshold standard* just the concept of "water clarity" as shown in Line 3 of Figure 1-2 below? Or is it the more specific "goal statement" of the type shown in Point b above? Or is it the indicator/measurement (Point c above)? Or the desired future measurement/target level (Point d above)? Or some combination of these things???

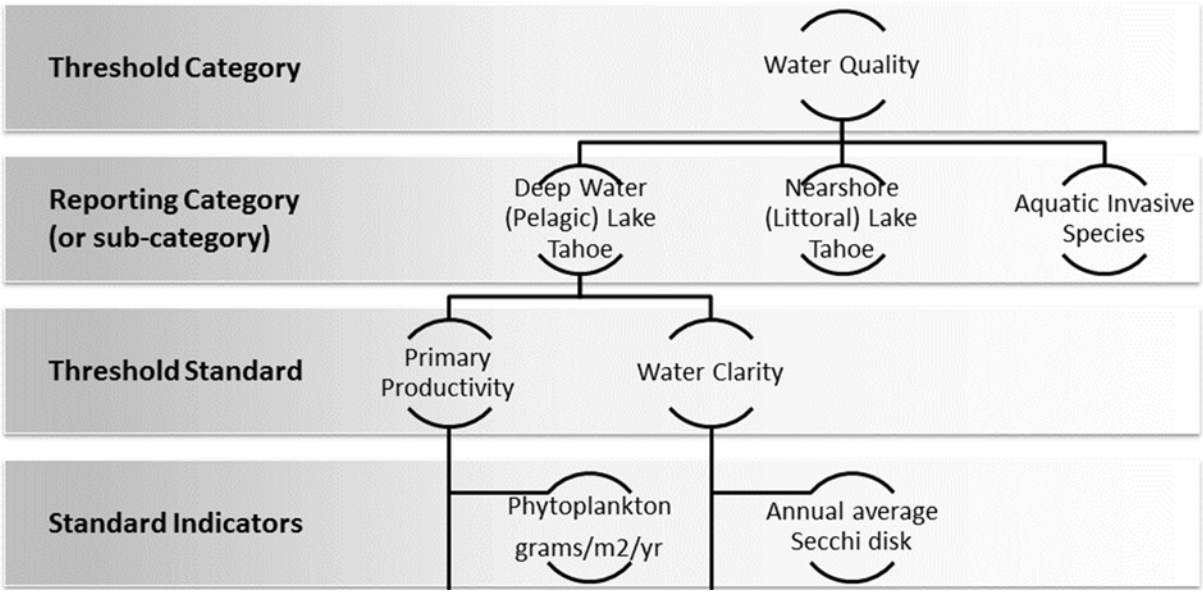


Figure 1-2. Example of how threshold standards are hierarchically organized under indicator reporting

I would love to see a figure or a text box in Chapter 13 that 1) defines and illustrates these concepts with a simple example as above, and 2) makes it clear what aspects of this system this review is focused on. Right now, all these concepts are smushed together under the rubric "threshold standard system" causing a fair amount of confusion. This definitional piece would both facilitate comparisons of the Tahoe system with other systems and make the Tahoe system more understandable and internally consistent.

#3. Does the chapter clearly lay out the rationale for the proposed assessment?

The need comes through pretty clearly in the intro.

#4. Is the proposed assessment framework logically consistent and clearly articulated?

The framework could benefit from clarifying terms as outlined above in my answer to Question #2 – and then applying whatever terms you all decide on consistently throughout the chapter. Specifically, it would be good to know what specific entities the assessment framework is being applied to. This might then lead to a reshuffling of the Question in Tables 13-1 and 13-2. To my mind, four main questions need to be asked about each standard being considered.

1. **Is the Threshold Standard / Goal *Relevant*?** -- Is the parameter being assessed useful for key management decisions (aka does it inform the status of a key element of the system? Or the effectiveness of a management intervention?) If not (or if it is redundant with other parameters), then immediately terminate and don't give it any more time / attention.
2. **Is this the Best Indicator/Method for Assessing Progress Toward this Goal?** -- Does the selected indicator (or set of indicators) and associated methods measure the Standard in a <u>sufficiently</u> accurate, consistent, and reliable manner? Important – note the emphasis on sufficient – you only need a minimum level of accuracy etc. needed for management decisions – see my specific comment on p 13-6 below that talks about the distinction between an indicator that is measureable vs. one that is quantitative.

3. **Has the Desired Future Value/Target Level Been Set at an Appropriate/Relevant Level?** – Does the desired threshold value reflect the desired future condition of the target – what you NEED to accomplish? What is the scientific evidence / certainty supporting this claim? Note that this is different than the confidence value in your measurement itself – which is what I think the border width of the circles is trying to show. As one example, Miradi Software supports an assessment of confidence in the threshold values and a separate assessment of confidence in the measurement value.
4. **Has the Desired Future Value/Target Been Set at an Achievable Level?** – Is the desired threshold realistic? If what you CAN accomplish is less than what you NEED to accomplish, then this becomes a trigger for important discussions and re-prioritizations of effort. But you can't throw out a Standard just because you can't reach it.

#5. Is the proposed approach for assessing the threshold standards consistent with best practice in the field of monitoring and evaluation? Is the proposed approach technically sound?

So if I understand this question correctly, you are NOT asking whether the measurement system is consistent with best practice, but instead whether your assessment of the measurement system (a meta-assessment) is consistent with best practice.

Honestly, I don't know of many attempts to systematically do this kind of meta-assessment. But I think you are generally on the right track – if you take into account some of the suggestions above.

#6. Are there additional questions that you think would help decision makers prioritize investments in updating the threshold standards?

The piece that seems to be missing here is an overall conceptual model or system map that shows the major "targets" that are the focus of management efforts and the major "threats or barriers" that need to be overcome to get the system to the desired future condition. Absent this picture (which may exist in other parts of the overall report) it is hard to know which standards are most relevant, which is the key to prioritization as outlined in my answer to the next question.

#7. Budgetary constraints mean that it may not be possible to review the entire suite of 171 standards all at once. Will the proposed assessment provide information necessary to prioritize review efforts? Are there additional pieces of information that may help inform the prioritization of scare resources?

There are actually two levels of prioritization needed here. 1) Which standards need to be prioritized for review, and 2) Which standards need to be prioritized for implementation. Unfortunately, I think most of your questions in Table 13-2 are aimed at the latter. To my mind, the key criteria for prioritizing which standards to review would include:

a) *Relevance of the Standard to Key Management Decisions* – Is this a standard that managers rely on and that provides binding constraints on the system. Here you obviously want to review only those that are most relevant / binding.

b) *Degree of "Settledness / Controversy"* – Is the science/theory behind the standard and the methods for setting goals and measuring the standard well-established / non-controversial or in flux / controversial? Obviously, the latter are more ripe for review.

c) *Utility / Happiness* – Are your stakeholders generally happy with the standard and merrily using it and reporting on it? Or is it causing angst and problems? Here again, the latter probably deserve more attention.

d) *Cost/Feasibility* – All else being equal across the previous three criteria, I would focus more on difficult expensive standards than cheap easy ones.

Once you get through the review and get to prioritization of standards for implementation, your questions in Tables 2-1 and especially 2-2 will be more useful. But I suspect you all are going to need stronger prioritization tools to really force folks into narrowing things down. If it were me, I would give people a "budget" (say 100 points of time and treasure) and ask them to allocate their points to the hopefully already narrowed list of standards – or at least to maybe a stratified cluster of standards around a key theme since most folks probably can't manage prioritizing across 171 items!

To be blunt, if you all can't even find time to review 171 Standards, how on earth do you expect to implement all of them??!!

#8. General impressions of the TRPA standards: This is your opportunity to provide general feedback on standards. Are they similar to other systems you have worked with? Are there points of difference from other systems you have worked with? If you were charged with leading the review and update of the standards, how and where would you start the process? Is there any guidance you would like to provide the agency and its partners as they begin the review process?

Overall, it is great that TRPA is taking this kind of approach. If I were charged with leading the review, I would follow the steps I outlined above.

One closely related system that you might want to take a look at is the Viability Analysis methodology developed by The Nature Conservancy and the Conservation Measures Partnership and that is supported by Miradi Software. To my mind, this is the best system that has been developed for clearly articulating focal conservation targets, goals, key attributes, indicators, measurements and desired future condition. This approach provides status assessments that are similar to those in the earlier chapters in this report. But there are a few differences that would be worth looking at including the concept of key ecological attributes, the idea that the thresholds are set at values that determine viability in poor, fair, good and very good buckets, and the treatment of uncertainty in setting these threshold values.

See the following for references:

Parrish, J.D., D.P. Braun, and R.S. Unnasch. 2003. Are we conserving what we say we are? Measuring ecological integrity within protected areas. **Bioscience** 53: 851-860.

TNC. 2007. **Conservation Action Planning Toolbox. 3. Assess Viability of Focal Conservation Targets.** http://conserveonline.org/workspaces/cbdgateway/cap/practices/bp_3

Braun, D.P. 2007. **Advanced Guidance for Step 3: Assessing the Viability of Focal Conservation Targets.** The Nature Conservancy.

Miradi Software. Available at www.miradi.org.

Some Specific Notes on the Chapter

The following are notes that I took as I read through the chapter. There is obviously some redundancy with the general comments above.

p 13-3 I wonder if it is necessary to consider modifying even the 9 categories themselves. Would help to have a conceptual model to understand how the dimensions fit together.

p 13-4 I generally applaud the use of the SMART criteria. However, the version offered here is not the optimal one in my mind. In particular, if you use *A = Achievable* and *R = Realistic*, these are redundant parameters. I much prefer to have the *R = Relevant*. This is the most important criterion of all. For example, I could set a goal to "tie and untie my shoes 3 times in the next 10 minutes." This is *Specific, Measureable, Achievable, Realistic,* and *Timebound* – but who the

heck cares? The question is whether the goal is relevant to policy objectives. *(note – see related comment on Table 13-2 below)*

p 13-5 Minor typo – *Measures of Success* should be cited as "Margoluis & Salafsky" and not the other way round as you all have it.

p 13-6 While I understand why you are not using the time-bound criterion, I'm not sure I fully support this decision. Yes these are long-term ambient environmental parameters. But it is still important to know from a policy and a management perspective what the timeframe for achieving the goals is expected to be. Is it going to take a month? A year? A decade? A century? I would support adding at least "order-of-magnitude" time estimates.

p 13-6 "The threshold standards are mixture of environmental standards, restoration goals, directives to engage in specific action, broad guidance, and narrative statements." I like that you all are acknowledging the heterogeneous nature of these goals. Most people don't want to seem to face up to this inevitable truth.

p 13-6 Not sure I understand the intent of the paragraph on the bottom of this page talking about the second set of categorization questions. It wasn't until I read Table 13.2 itself that I understood what you meant by this type of question.

P 13-6 In setting thresholds (as in all applied scientific endeavors) it is "better to have approximate answers to exact questions rather than exact answers to less useful questions." This has the most bearing on the use of the "Measureable" assessment criterion. People often confuse "measureable" with "quantitative." I'd rather see the "right indicator" measured qualitatively or approximately than the "wrong indicator" measured with great precision!

To this end, I encourage people in setting thresholds for status variables to think about the desired condition of the underlying resource target. You might start with a threshold that says we need "less periphyton algae in the littoral zone." You then might over time refine this to say "no visible algae mats in 80% of the littoral zone" and then maybe eventually to something much more specific and quantifiable like "no more than 200 algae cells per liter of water sampled during July peak low water flow using the Beta7 water sampling protocol" (please note that I obviously know nothing about periphyton algae and am just making up these measures!!). But unless there are regulatory, legal or scientific needs requiring the detailed indicator/method, the second or even the first one may be sufficient for management purposes.

Table 13-1 Per my comment on p 13-4 above, in the table you are now using different terms for the letters in SMART – specifically you are using both *Relevant* (which I like as noted above) but also *Attributable* (which now seems to overlap with *Relevant* in its intent and is confusing because you now have two "A" terms). I might consider dropping the *Attributable*. Obviously, whatever categories you end up with, they should be consistent throughout the doc.

Table 13-2 I REALLY like the *Focus* criterion – but I would like to see a more extended treatment of the concept of theories of change and the difference (in Open Standards terms) between objectives (desired state of intermediate results) and goals (desired state of ultimate targets).