UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

September 28, 2006

EPA-SAB-06-011

The Honorable Stephen L. Johnson
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: SAB Review of the EPA Region 6 Geographic Information System Screening Tool

Dear Administrator Johnson:

The EPA Region 6 Compliance Assurance and Enforcement Division requested Science Advisory Board (SAB) review of the Region's Geographic Information System Screening Tool (GISST). The GISST is a geographic information system-based tool used for rapid preliminary environmental assessment and reconnaissance. EPA Region 6 has applied the GISST to develop information for evaluating environmental impact statements required under the National Environmental Policy Act (NEPA). An SAB panel reviewed the strengths and limitations of the GISST. The enclosed SAB report addresses EPA's charge questions to the Panel.

The SAB commends Region 6 for developing the GISST. Geographic Information System (GIS) capabilities and data layers provide essential support for efficient, timely, and proactive NEPA evaluations and other Regional responsibilities. The SAB notes that several elements make GISST evaluations different from other GIS evaluations. Unique GISST elements include: 1) the criteria scoring process, 2) criteria subset selection, and 3) the process that highlights important drivers of concern for further analysis. These elements make the GISST an objective, spatially explicit tool for conducting, broad-stroke preliminary evaluations in a timely fashion. The SAB finds that it is reasonable and appropriate to evaluate individual criteria or suites of criteria in the GISST to detect potential environmental impacts of certain types of projects.

The SAB also has identified limitations in the methodological approach used in the current version of the GISST. Because of these limitations, the aggregate GISST vulnerability or impact score should not be used at this time to conduct assessments for decision-making, such as setting priorities or considering the impacts of project alternatives. The SAB has recommended improvements to make the GISST suitable for future use. In this regard the SAB finds that:

- The current version of the GISST provides only a single vulnerability or impact score that can mask important differences in individual data layers used for an assessment. Such differences must be considered when evaluating the potential environmental impacts of project alternatives as part of an overall strategy to achieve assessment objectives.

- The current version of the GISST does not adequately utilize modern decision analytic and statistical science in its development of numerical scoring. These issues must be considered when combining and evaluating data layers in the GISST.

- EPA should draw upon the available expertise and large literature on multi-attribute decision-making to further develop the GISST.

GIS-based assessment tools are needed to provide essential support for many EPA activities. Various Agency program offices and regions have developed key components of these tools. In addition to the Region 6 GISST, other examples include the Region 4 Southeastern Ecological Framework, the Region 5 Critical Ecosystem Assessment Model, the Office of Water's Index of Watershed Indicators, and the Office of Research and Development's Regional Vulnerability Assessment methods. In reviewing all of these assessment tools, the SAB has noted common problems of inadequate involvement of statistics and decision analytic expertise, weighting of metrics, and procedures used to combine data. Furthermore, despite these initiatives, the Agency still does not have a unified single accepted approach for using spatially explicit information for environmental decision-making. The compartmentalized development of GIS-based tools and data by EPA program offices and regions is inefficient, given budgetary constraints and the high value of these tools for environmental decision-making. Using spatially explicit information for decision-making is a critical need that EPA must address. The SAB therefore strongly urges EPA to undertake an initiative to define a unified framework for the development of these types of tools across the Agency. A national effort to develop such a unified framework will require decision analytic expertise, and the resulting tool should be easy to modify to meet local and regional needs.

Sincerely,

/signed/

Dr. M. Granger Morgan, Chair
EPA Science Advisory Board

/signed/

Dr. Virginia Dale, Chair
Geographic Information System
Screening Tool Review Panel
EPA Science Advisory Board

NOTICE

This report has been written as part of the activities of the EPA Science Advisory Board, a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The Board is structured to provide balanced, expert assessment of scientific matters related to the problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use. Reports of the EPA Science Advisory Board are posted on the EPA website at http://www.epa.gov/sab.

U.S. Environmental Protection Agency
Science Advisory Board
Geographic Information System Screening Tool Review Panel

CHAIR

Dr. Virginia Dale, Corporate Fellow, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

MEMBERS

Mr. DeWitt Braud, Director, Academic Area, Coastal Studies Institute, Louisiana State University, Baton Rouge, LA

Dr. Ivan J. Fernandez, Professor, Department of Plant, Soil and Environmental Sciences, University of Maine, Orono, ME

Dr. Carol Johnston, Professor, Center for Biocomplexity Studies, South Dakota State University, Brookings, SD

Dr. William Mitsch, Professor, Olentangy River Wetland Research Park, The Ohio State University, Columbus, OH

Dr. Thomas C. Mueller, Professor, Department of Plant Sciences, University of Tennessee, Knoxville, TN

Dr. Michael C. Newman, Professor of Marine Science, School of Marine Sciences, Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA

Dr. James Oris, Professor, Department of Zoology, Miami University, Oxford, OH

Dr. Charles Rabeni, Leader, Missouri Cooperative Fish and Wildlife Research Unit, U.S. Geological Survey, Columbia, MO

Dr. Mark Ridgley, Professor and Chair, Department of Geography, University of Hawaii at Manoa, Honolulu, HI

Dr. Amanda Rodewald, Assistant Professor, School of Natural Resources, The Ohio State University, Columbus, OH

Dr. James Sanders, Director, Skidaway Institute of Oceanography, Savannah, GA

Dr. David Stoms, Associate Research Scientist, Institute for Computational Earth Systems Science, University of California at Santa Barbara, Santa Barbara, CA

Mr. Timothy Thompson, Senior Environmental Scientist, Science, Engineering, and the Environment, LLC, Seattle, WA

Dr. Robert Twiss, Professor, University of California-Berkeley, Ross, CA

SCIENCE ADVISORY BOARD STAFF

Dr. Thomas Armitage, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

U.S. Environmental Protection Agency
Science Advisory Board

CHAIR
Dr. M. Granger Morgan, Carnegie Mellon University, Pittsburgh, PA

MEMBERS
Dr. Gregory Biddinger, ExxonMobil Biomedical Sciences, Inc, Houston, TX

Dr. James Bus, The Dow Chemical Company, Midland, MI

Dr. Trudy Ann Cameron, University of Oregon, Eugene, OR

Dr. Deborah Cory-Slechta, University of Medicine and Dentistry of New Jersey and Rutgers State University, Piscataway, NJ

Dr. Maureen L. Cropper, University of Maryland, College Park, MD

Dr. Virginia Dale, Oak Ridge National Laboratory, Oak Ridge, TN

Dr. Kenneth Dickson, University of North Texas, Denton, TX

Dr. Baruch Fischhoff, Carnegie Mellon University, Pittsburgh, PA

Dr. A. Myrick Freeman, Bowdoin College, Brunswick, ME

Dr. James Galloway, University of Virginia, Charlottesville, VA

Dr. Lawrence Goulder, Stanford University, Stanford, CA

Dr. Rogene Henderson, Lovelace Respiratory Research Institute, Albuquerque, NM

Dr. Philip Hopke, Clarkson University, Potsdam, NY

Dr. James H. Johnson, Howard University, Washington, DC

Dr. Meryl Karol, University of Pittsburgh, Pittsburgh, PA

Dr. Catherine Kling, Iowa State University, Ames, IA

Dr. George Lambert, UMDNJ-Robert Wood Johnson Medical School/ University of Medicine and Dentistry of New Jersey, New Brunswick, NJ

Dr. Jill Lipoti, New Jersey Department of Environmental Protection, Trenton, NJ

Dr. Genevieve Matanoski, Johns Hopkins University, Baltimore, MD

Dr. Michael J. McFarland, Utah State University, Logan, UT

Dr. Jana Milford, University of Colorado, Boulder, CO

Dr. Rebecca Parkin, The George Washington University, Washington, DC

Mr. David Rejeski, Woodrow Wilson International Center for Scholars, Washington, DC

Dr. Joan B. Rose, Michigan State University, E. Lansing, MI

Dr. Kathleen Segerson, University of Connecticut, Storrs, CT

Dr. Kristin Shrader-Frechette, University of Notre Dame, Notre Dame, IN

Dr. Robert Stavins, Harvard University, Cambridge, MA

Dr. Deborah Swackhamer, University of Minnesota, Minneapolis, MN

Dr. Thomas L. Theis, University of Illinois at Chicago, Chicago, IL

Dr. Valerie Thomas, Georgia Institute of Technology, Atlanta, GA

Dr. Barton H. (Buzz) Thompson, Jr., Stanford University, Stanford, CA

Dr. Robert Twiss, University of California-Berkeley, Ross, CA

Dr. Terry F. Young, Environmental Defense, Oakland, CA

Dr. Lauren Zeise, California Environmental Protection Agency, Oakland, CA

SCIENCE ADVISORY BOARD STAFF

Mr. Thomas Miller, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The EPA Region 6 Compliance Assurance and Enforcement Division requested that the Science Advisory Board (SAB) review the Region's Geographic Information System Screening Tool (GISST). The GISST is a geographic information system-based tool for evaluating the potential environmental impacts of large projects such as the construction of roads and the issuance of permits for water treatment plants. A mathematical algorithm is used in the GISST to evaluate various impact or vulnerability criteria data layers and derive an aggregate score that has been called the "potential for significant environmental risk" of a project. The GISST has been used to provide information for preparing and reviewing environmental assessments and impact statements required under the National Environmental Policy Act (NEPA). The SAB GISST Review Panel met in December 2005 and deliberated on six charge questions. These questions focused on: 1) whether the GISST methodology is reasonable and appropriate for use in conducting initial assessments of potential environmental vulnerability and impacts; 2) the strengths and limitations of the GISST as a tool for use in prioritizing and comparing environmental vulnerabilities and impacts for decision-making; and 3) recommendations to improve the GISST User's Manual and documentation.

The Panel commends Region 6 for developing the GISST and providing strong GIS support to its environmental managers. Geographic Information System (GIS) capabilities and data layers provide essential support for efficient, timely, and proactive NEPA evaluations and other Regional responsibilities. GISST is an exploratory tool that can improve understanding of how potential decisions can affect the systems being considered. Three types of potential NEPA issues can be explored using the GISST: 1) consideration of whether impacts are likely to occur (the scoping process); 2) consideration of whether impacts have been adequately addressed in an environmental document; and 3) consideration of alternatives that appear to have the least impact. The benefits of compiling data layers for the GISST have undoubtedly extended to other applications within Region 6. The Panel finds that it is reasonable and appropriate to use the scores of individual GISST criteria, or suites of criteria corresponding to different types of vulnerability, for conducting broad-stroke preliminary evaluations to detect potential environmental impacts of certain types of projects. However, the Panel has identified limitations in the methodological approach used in the current version of the GISST. Because of these limitations, the aggregate GISST score should not be used in assessments for decision-making.

Concerted EPA effort is needed to develop assessment tools like the GISST. Various EPA program offices and regions have created screening tools similar to the GISST since GIS technology became widely available in the 1990s. Examples include the EPA Region 4 Southeastern Ecological Framework, the EPA Region 5 Critical Ecosystem Assessment Model, the EPA Office of Water's Index of Watershed Indicators, the EPA Office of Research and Development's Regional Vulnerability Assessment methods, and the NEPAssist web-based mapping tool developed by EPA's Office of Federal Activities. Leibowitz et al. (1992) utilized similar methods in the synoptic approach developed for cumulative impact assessment of wetlands. The principal goal in all of these developments was to harness the power of geospatial data as an aid to environmental decision-making. However, the utility of these GIS-based decision assistance tools is limited by the amount and quality of the underlying geospatial

environmental data sets and the lack of suitable indicators of biological and ecological effects.

Many EPA program offices and regions have common needs for data sets such as those developed for the GISST. Particular examples include the geographically aggregated summaries of point data from EPA's Storage and Retrieval System (STORET), the Toxics Release Inventory (TRI), and National Pollution Discharge Elimination System (NPDES) databases. The compartmentalized development of GIS tools and data by EPA program offices and regions is suboptimal, given the universal need for such tools. It does not help the development of a national perspective, nor is it an efficient use of scarce resources. The Panel urges EPA to make a concerted effort to develop a unifying framework for the creation of these tools and suggests that they could be provided by EPA's Office of Environmental Information. The SAB Ecological Processes and Effects Committee made a similar recommendation in its review of EPA's Index of Watershed Indicators (U.S. EPA Science Advisory Board, 1999). In that review the SAB recommended that, "the Agency should add more indicators of biological and ecosystem effects to the Index of Watershed Indicators." The panel also urges that future development of the GISST be consistent with the principles embodied in EPA's *Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models* (U.S. EPA Office of Science Policy, 2003).

GISST Mathematical Algorithm

The Panel was asked to comment on the reasonableness and appropriateness of using the GISST algorithm for conducting screening-level[1] evaluations. The algorithm is simple and straightforward, providing a reasonable way to conduct initial evaluations. However, the algorithm generates an aggregate vulnerability or impact score that may mask important differences in individual criteria scores. Such differences must be considered when evaluating project alternatives. The panel suggests that the algorithm score itself be de-emphasized and the GISST be used as part of a screening process that clearly defines the limits of integrative indices and promotes the subjective scientific evaluation of numerical output and supporting information. In order to further explore uses of the GISST for rapid environmental assessment and preliminary reconnaissance, EPA should develop optional models for aggregating the criteria used in the algorithm and weighting[2] them according to their relative importance to decision-makers. The Panel emphasizes, however, that in summing criterion scores, high scores can offset or compensate for low ones and hence mask potential problem areas. Such compensation reflects implicit tradeoffs, functions of the coefficients and units of the scores being summed, regardless of the intent and preferences of the users and clients of the analysis. Since such intent and preferences are unlikely to be known, or known fully, the summation of criterion scores is unlikely to express participants' acceptable tradeoffs. In addition, to be mathematically legitimate, one must assess or rate an alternative's (e.g., site's) performance on the relevant criteria on measurement scales appropriate for the arithmetic operations – addition, subtraction, multiplication, and division – subsequently used in the algorithm. Thus, summation of impact and vulnerability scores should only be used in an exploratory vein and as a preliminary reconnaissance. The potential utility of summing scores lies with uncovering potential problems

[1] In this context a screening-level evaluation is a rapid preliminary assessment of potential environmental and societal impacts of certain types of projects.

[2] Weighting is the process of assigning multipliers to a set of quantities to indicate the relative importance of each quantity within the set.

derived from the interaction of multiple impacts. Summation should not be used to verify the absence of adverse consequences. It is recommended that the GISST algorithm be re-evaluated, taking advantage of additional expertise in decision science and statistics to address these and other concerns expressed in this report.

The Panel was also asked to consider the reasonableness and appropriateness of the method used by EPA Region 6 to evaluate environmental vulnerability in the Interstate Highway 69 case. The published GISST algorithm was not used in this case. Instead, the vulnerability within the highway corridor was evaluated by summing the vulnerability scores within 1 km^2 areas. The Panel finds that the approach described in this case can be used to explore potential vulnerabilities or impacts if the criteria scores are averaged and as long as they represent only impacts or only vulnerabilities. Other concerns regarding mathematical operations for aggregating scores are described below.

GISST Criteria

The Panel was asked to comment on whether the GISST criteria (the kinds of data used to score vulnerability and impact) were reasonable and appropriate for use in evaluations of the potential impacts of projects and vulnerability of project areas. The Panel finds that the individual GISST criteria are intuitive, but there is a need to better describe how to use groups of criteria. The criteria have competing purposes with varied thresholds that can result in criteria being combined in illogical ways (e.g., combining noise and odor criteria scores with scores for use of energy efficient appliances). Criteria categories need to be defined consistently so that combinations or groupings make sense in an overall strategy. Instead of focusing on the appropriateness of a particular group of criteria, it would be better to define a general process for selecting groups of criteria for use in various kinds of evaluations. The GISST could be improved by developing impact templates. These templates could identify the kinds of impacts that might be associated with particular project types and the key criteria relevant to evaluation of those impacts.

The Panel is also concerned that EPA has not clearly described the differences between the criteria used to determine vulnerability and those used to determine impact. Some criteria reflect both vulnerability and impact but others reflect only one of these indices. More detailed descriptions and explanations of the criteria and supporting databases are needed in the GISST documentation. A statistical examination of the supporting data is needed to: 1) determine relationships between the criteria data layers, 2) identify potentially redundant measures, and 3) increase the soundness of the rating scales.

GISST Scoring System

The Panel was asked whether the GISST 1-5 scoring scale was reasonable for use with different data sets and data coverages to develop an initial assessment of the potential cumulative impacts of proposed projects. The GISST scoring system is reasonable for use in detecting potential individual concerns where no mathematical operations are required for aggregating scores. The advantage of the GISST scoring system is that it allows evaluators to simplify a combination of diverse criteria and show contrasts in the assessment of potential impacts among

project alternatives. A disadvantage of the scoring system is that assessors lose the ability to see specific information that may be provided by the underlying data for a particular assessment.

A number of issues must be considered if mathematical operations are required to aggregate criteria scores. If the GISST criteria values are ordinal, it is not appropriate to average or multiply them. The GISST criteria values seem to have been binned or scaled using different techniques or functions so it may be inappropriate to equate their scales during summing. A number of statistical issues must also be considered. Many of the criteria seem to overlap and possibly interact. The potential consequences of interactions are unclear and must be evaluated. The Panel suggests that the GISST criteria scoring system could be substantially improved by involving more decision analytic and statistical expertise in the development of an expert opinion-based ranking system. Explicit decision rules and criteria weightings should also be defined.

Use of the GISST in the NEPA Process to Prioritize Project Impacts for More Detailed Analysis

The Panel commented on the strengths and limitations of the GISST as a screening tool to prioritize project impacts in the NEPA process. If sets of core criteria were identified for evaluating certain kinds of projects, the GISST could be used in the NEPA process to assist EPA in reviewing and scoping environmental assessments and environmental impact statements. However, the GISST is inadequate for prioritizing impacts unless the following limitations are addressed: 1) the GISST criteria must be weighted so that scores represent the relative importance of various impacts to decision-makers; 2) the GISST criteria must reflect the concerns of stakeholders or decision-makers for specific problems being addressed; 3) a scoring system must be developed to reflect impacts and vulnerabilities in specific ecoregions and physiographic regions where scores will be applied; 4) inaccuracies resulting from imprecise data scales must be addressed; 5) the map classes selected and used in the GISST should be transparent and well documented.

Use of the GISST in the NEPA Process to Evaluate Environmental Impacts of Project Alternatives

The Panel was asked to comment on the usefulness of the GISST for evaluating the environmental impacts of project alternatives to help inform NEPA-related decision-making. The Panel finds that the GISST could be used to detect potential environmental impacts of project alternatives. However, to make the GISST useful for evaluating impacts of project alternatives and for making decisions, EPA must address the following limitations of the tool: 1) the GISST algorithm cannot identify specific impacts; 2) the GISST scoring system assigns equal weights to the criteria and may mask environmental impacts; 3) the GISST does not include a process for identifying relevant and meaningful criteria, thresholds, and impact levels; 4) spatial dependence of the cells in the GISST should be explicitly considered; 5) vulnerability and impact criteria can be confused; 6) the GISST cannot map projects simultaneously to illustrate advantages and disadvantages of alternatives; 7) the sum of the GISST average does not provide information about the pros and cons of alternatives; 8) the GISST will not be helpful in designing new alternatives unless criteria and objectives are defined *a priori*.

GISST User's Manual Enhancement

The Panel recommends enhancement of the GISST User's Manual. The GISST User's Manual provides a useful introduction to the tool, but it does not contain adequate instructions on how to operate the tool or interpret outputs. The User's Manual could be enhanced by including the following additional material:

- Background information about how the GISST supports NEPA assessments;
- Information explicitly describing the types of decisions that the GISST supports;
- Information describing the conceptual model underlying the GISST;
- The basis and process for selecting GISST criteria to use in an evaluation;
- Suggested approaches for integrating spatial data;
- The mathematical boundaries of output parameters and guidance on interpretation of results;
- Representative applications of the GISST;
- Definitions of key terms.

The Panel also recommends reorganization of the User's Manual and provides editorial comments.

1. INTRODUCTION

This report was prepared by the Science Advisory Board (SAB) Geographic Information System Screening Tool Review Panel (the "Panel") in response to a request by the EPA Region 6 Compliance Assurance and Enforcement Division to review the Region's Geographic Information System Screening Tool (GISST). The GISST is a descriptive geographic information system tool that has been used to evaluate the potential environmental impacts of large projects such as the construction of roads, the permitting of water treatment plants, and confined animal feeding operations. The GISST has also been used to evaluate the potential environmental vulnerability of the proposed sites of such projects. Such evaluations have been used by EPA Region 6 to facilitate decision-making and to prepare environmental impact statements required under the National Environmental Policy Act (NEPA).

Users of GISST assess the potential environmental impacts of proposed projects and the potential environmental vulnerabilities of project sites by visualizing various sets of geographically referenced data. These data sets underlie selected vulnerability and impact "criteria." For example, the GISST developers state that rainfall at a project location is evaluated as a vulnerability criterion because more rainfall can be associated with more infiltration to groundwater and runoff to surface water. The density of federally managed lands at a project location is evaluated as an impact criterion because federally managed lands tend to support a variety of ecological services and greater project impacts are anticipated in areas with higher densities of these lands. To facilitate decision-making, a scoring system (with a scale of one to five) is used in the GISST to evaluate data sets associated with each criterion. A lower score equals a lower level of potential impact or vulnerability. The GISST scoring system was developed by using arithmetic groupings to evaluate natural breaks in the data and applying the best professional judgment of EPA Region 6 staff.

In the GISST the degree of potential vulnerability of a watershed subunit, project area, or other appropriate geographical unit is defined as the average of the vulnerability criteria scores within the geographic unit. The degree of potential impact produced by the project is defined as the average of the impact criteria scores. A mathematical algorithm is used in the GISST to derive the "potential for significant environmental risk" associated with a project. This algorithm considers the ratio of the cumulative area affected by a project to the total area evaluated, the degree of potential vulnerability of the area evaluated, and the degree of potential impact produced by the project. The results of GISST analyses can be displayed on maps that can include data overlays generated using different criteria. The Panel sees GISST as an exploratory tool that can improve understanding of how potential decisions can affect the systems being considered. It is the use of GISST and not particular numbers calculated by it that most contribute to the decision-making process. However, the understanding provided by GISST is only as good and as complete as the underlying data layers. In the world of environmental decision-making, incomplete information is common, and decisions are often made using only available information. In cases where the scientific evidence is incomplete or contradictory, decisions are often made without scientific input. Wiens (1996) found such a case in the aftermath of the Valdez, Alaska oil spill. GISST provides a means to assimilate available information for rapid assessment. Because this tool is used for preliminary environmental

reconnaissance, it is appropriate to consider alternative weightings of data layers and interpretations of results. Such analysis of alternatives with GISST should be quite insightful.

The Panel reviewed the *Region 6 GIS Screening Tool (GISST) User's Manual* provided by EPA Region 6. The document contained: 1) an introduction in question and answer format that described the uses of the GISST; 2) background information on concepts underlying the GISST; 3) information on the development of the GISST algorithm and criteria; 4) case studies illustrating how the GISST has been applied; 5) the finalized GISST criteria; 6) additional GISST criteria that are under development; 7) the geographic information system program used in the GISST; and 8) a peer review history of the GISST.

The Panel notes that various EPA Regions and program offices have developed GIS-based assessment tools similar to the GISST. Examples include the EPA Region 4 Southeastern Ecological Framework, the EPA Region 5 Critical Ecosystem Assessment Model, the EPA Office of Water's Index of Watershed Indicators, the EPA Office of Research and Development's Regional Vulnerability Assessment methods, and the NEPAssist web-based mapping tool developed by EPA's Office of Federal Activities. Despite these initiatives, the Agency still does not have a unified single accepted approach for using spatially explicit information for environmental decision-making. The compartmentalized development of GIS-based tools and data by EPA program offices and regions is suboptimal, given budgetary constraints and the high value of these tools for environmental decision-making. The SAB therefore strongly urges EPA to undertake an initiative to define a unified framework for the creation of these types of tools. The panel also urges that future development of the GISST be consistent with the principles embodied in EPA's *Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models* (U.S. EPA Office of Science Policy, 2003).

2. CHARGE TO THE REVIEW PANEL

EPA Region 6 sought comment from the Science Advisory Board on the following issues: 1) whether the GISST methodology is reasonable and appropriate for use in conducting initial level assessments of potential environmental impacts and vulnerability, 2) the strengths and limitations of the GISST as a tool for use in prioritizing and comparing environmental vulnerabilities and impacts for decision-making, and 3) steps that can be taken to further develop the GISST User's manual and documentation. Six detailed charge questions were provided to the Panel.

3. RESPONSE TO THE CHARGE QUESTIONS

3.1 Question 1.1. The GISST mathematical algorithm (presented in Chapter 3 of the GISST User's Manual) for determining the "potential for significant environmental risk" of projects is a multiplicative formula using the watershed as the base unit. Please comment on the reasonableness and appropriateness of using this algorithm for conducting screening-level evaluations as described in the GISST User's Manual.

In the Interstate Highway 69 case study, the GISST algorithm was not used because

it was not beneficial to obtain one cumulative vulnerability score for the entire highway corridor. Instead, vulnerability within the corridor was evaluated by summing the scores of vulnerability criteria within 1 km^2 areas in a grid system. Please comment on the reasonableness and appropriateness of this method for conducting an initial screening-level evaluation.

Reasonableness and Appropriateness of GISST for Conducting Screening Level Evaluations

The GISST algorithm is straightforward, consisting of three multiplicative parts. The first deals with the overall area affected, the second with potential for vulnerability, and the third with potential for impact. The GISST algorithm has the advantage of being a simple tool that allows the aggregation of layers of information for environmental assessment screening. In this regard, geographic information system (GIS) based approaches to screening such as the GISST are very useful. They should be further developed, and their use should be promoted. While the GISST algorithm provides a single score, the score itself should be de-emphasized. Much of the value of the approach is in the process itself and the real strength of the tool is in the individual data layers. Focusing on the algorithm alone can mask some of these details and their value. The GISST tool should be part of a screening process that clearly defines limits of integrative indices and promotes the subjective scientific evaluation of the numerical output and information that supports it. This screening process can include other visual representations of GIS-based data layers that allow for visual identification of underlying layer information (e.g., "radar plots" that enable visualization of the criteria values).

The GISST algorithm provides a reasonable way to conduct an initial evaluation of potential environmental impacts, but the limitations and caveats associated with the use of this simple algorithm and discussed below must be considered and addressed. The *multiplicative* nature of the algorithm is appropriate since it accords with the *Risk = Probability x Consequences* structure of the most commonly used definition of risk. It also makes sense that if either vulnerability or impacts are nil (zero), then their product should also be nil and no further attention should be given to assessing the risks of the project. The scaling of the potential project impact by project size relative to watershed size is simplistic but reasonable. However, the specific algorithm used in the GISST needs to be re-evaluated taking advantage of the available literature and expertise in decision science and statistics, particularly in the area of multi-attribute decision-making, to address the following concerns raised during this review as well as other concerns that may be identified.

- Compensatory effect. The algorithm is an example of the "compensatory" class of multi-attribute evaluation methods, in that high scores on some dimensions (criteria) can compensate for low scores on others. This relationship arises from the fact that the overall *vulnerability* of a project alternative is measured as an average of the vulnerability scores of individual vulnerability criteria, and likewise for overall *impact*. One caveat thus pertains to this compensatory structure: it could be undesirable that very low scores on some dimensions can be countered by high performances on others. Such tradeoffs are commonly *not* acceptable in many biological, ecological, and ethical situations: an organism cannot substitute more energy intake (calories) for less water; a greater abundance or biomass of a common species may not offset losses of a rare or endangered species; more very rich people may not compensate for more dire poverty (*cf.*, the

computation of environmental justice scores as the average of three different individual criterion scores). This compensatory scheme masks potential differences between project alternatives, since the same or similar averages can arise from very different constituent scores. Unless one scrutinizes all individual scores, one will not necessarily notice such differences. Such compensatory structures are common for utilitarian welfare functions, where one attempts to maximize total or average welfare, or utilitarian damage or impact functions, where the total or average cost (pollution, risk) is minimized. Compensatory methods do not address distributional aspects like equity. Users need to be aware of these characteristics of compensatory, utilitarian structures.

- Weighting the criteria. In the GISST algorithm, the criteria scores are averaged to derive either the degree of vulnerability (D_v) or the degree of impact (D_i), although the authors point out that sometimes it might be best to sum the criteria in each group. Three problems argue against such averaging and/or summing. First, to be mathematically legitimate, criterion scores must be measured on an interval scale (i.e., on the 1-5 scale, unit differences between scores are all of the same magnitude) or ratio scale (i.e., on a 1-5 scale the scores are stated with respect to a rational zero) if they are to be averaged; if they are only summed, and not averaged afterward, then scores must be on a ratio scale. It is not clear that all of the criteria scores meet these requirements. Many of the criteria scores appear to have ordinal properties (i.e., on the 1-5 scale the scores are only indicative of more or less of a characteristic), and it is not mathematically legitimate to sum or average ordinal scores. Second, the average of a number of different scores fails to reveal the range of those scores. Extreme values (e.g., those that fall above or below some important threshold) will thus be concealed, and the power of the tool to illuminate (”flag”) potentially problematic effects will be diminished. Third, summing scores on different dimensions, even if they have the same units (e.g., kg of phosphorus and kg of nitrogen), implies a value tradeoff between any two scores and can mask critical impacts. Since this tradeoff is equal to the negative of the ratio between their coefficients, even “intentionally unweighted” scores (i.e. with coefficients equal to 1.0) are in fact implicitly weighted equally (changing units, for example from kilograms to tons, reveals this immediately). Hence, weighting and tradeoffs are unavoidable in the additive value (impact) function. The tradeoffs can mask critical impacts because, in the summation, desirable effects (whether many small ones, or a few very large ones) can compensate undesirable ones (whether small or large), since it is only the sum that matters.

  The implication of this unavoidable weighting and potential masking of impacts is that summed criterion scores should not be used as the only signal of potential adverse impacts unless the weights actually represent clients’ (decision-makers’, stakeholders’ and other relevant actors’) desirable tradeoffs between different criterion scores. Given the difficulty of determining such tradeoffs (weights) for public problems, and the likelihood that a linear (weighted-sum) function would represent them, summed scores should only be used in an exploratory vein: extreme sums can indicate that there might be a problem, but only examination of the individual scores will confirm this and reveal where and on what dimensions. Nevertheless, any analysis that moves from separate measures to summary measures must contain normative judgments, and if the analysis is used in an exploratory way, it must be easy for users to employ different judgments. It will therefore

be useful if the GISST can be modified to allow assignments of different weights to criteria scores since a differently weighted function might reveal potential problem areas otherwise hidden. Nonetheless, it is crucial that the user be aware of "false negatives": because of the compensatory nature of the summation, absence of extreme sums does not imply absence of significant adverse impacts.

- Measurement of risk. The conceptual nature of GISST algorithm core, *vulnerability x impacts*, needs refinement if it is to measure something akin to risk (or as stated in the GISST User's Manual, "the potential for significant risk"). The uncertainty reflected in the standard definition of risk is represented by a *probability* term. In the GISST, *uncertainty* is apparently represented by *vulnerability*, and *impacts* are synonymous with *consequences*. However vulnerability is not precisely defined in the GISST User's Manual. A number of concepts and operational definitions of vulnerability are found in the literature, but few would equate vulnerability with probability, likelihood, or some other notion of uncertainty, and unless vulnerability is measured on a scale of 0-1 it would not reflect probability. Thus, some measure of uncertainty of consequences would have to be added to the GISST algorithm if it were to be used to determine the potential for risk. Such a modification would not be necessary if EPA continues to use the GISST as a tool to identify the potential for impacts.

- Choice and number of criteria. Another problem associated with the GISST algorithm lies in the choice and the number of criteria to be used in the analysis. Users can select and apply criteria deemed appropriate for their particular project. This adds specificity and provides flexibility. However, it also provides a mechanism by which the impact of important or "problem criteria" can be masked by relatively benign ones. For example, if 5 criteria are chosen, and their scores are 1, 1, 1, 1, and 5, the averaged score is 1.8. If only two of these criteria are chosen, their scores could be 1 and 5 resulting in an average score of 3. The GISST authors are aware of this important issue, and point out in the GISST User's Manual that it is also important to examine the scores of each criterion to look for potential hot spots. This in essence is the weighting procedure discussed above. It is also important to always keep in mind the number of criteria used to determine D_v relative to D_i, so that each of those groupings have equal weight, if that is desired.

- Alternate algorithms. In order to conduct sensitivity analyses and further explore uses of the GISST for rapid environmental assessment and reconnaissance, EPA should develop some optional models of aggregating and weighting the criteria, not just the one shown in the GISST User's Manual. EPA Region 6 has not used the algorithm in the User's Manual for the major test case of the Interstate Highway 69 Corridor. It is therefore not clear why this algorithm is published as the only available choice. Other possible integrative techniques should be provided in the GISST User's Manual.

- Modifiable Area Effect. The GISST criteria can be applied in a number of different ways depending upon what "appropriate geographic unit" (AGU) is selected for an evaluation. The selected unit can be a physiographic unit watershed, a U.S. Geological Service hydrologic cataloging unit (HUC), or a square kilometer, and the results will vary depending on that selection. If the AGU is very large relative to the affected area, then the

GISST score will be small regardless of what the D_v and D_i scores are because the percentage of area affected relative to AGU will be very small. While the GISST is a screening tool and there is no "safe" score, small scores can still leave the impression that there are few problems associated with a project, when in fact there might be significant ones.

- Landscape dynamics. System or landscape-level dynamics are not addressed in the GISST algorithm and should be considered. For example, the beginning of a highway construction project may cause a big disruption of urban lives and affect wildlife but things may improve after a period of time (e.g., in the Florida Everglades, black panthers now use "underpass ramps" which connect the landscape made disjoint by the highway). Conversely, nitrate pollution from confined animal feeding operation (CAFO) systems may not show up for decades in groundwater. EPA should be cognizant that the problem should define the tool, not vice versa.

- Quality of data. The presentation of GISST analyses as numerical output should include the underlying "metadata" as part of the overall report in order to provide the user or reviewer information on: data source, dates, extent, quality, completeness, processing steps, responsibility, and contact information. In addition, sensitivity analyses and "goodness-of-fit" measures could be provided to describe how well the index and its components can address the questions asked. It is apparent that some GISST analyses clearly provide answers to questions, but others could result in quite different answers depending on the judgment of the user. The Panel emphasizes that numeric indices can be very useful but they should not be used in a vacuum to make decisions.

- Math errors in the algorithm should be corrected. For example, the expressions describing D_i on page 21 and in Appendices A and B of the GISST User's Manual mean different things. It is also not clear in the algorithm what is being added with the summation sign (the Panel questions, for example, whether A_w is fixed or a variable).

Interstate Highway 69 Case

The Panel considered the reasonableness and appropriateness of the method used by EPA Region 6 to evaluate vulnerability in the Interstate Highway 69 case. The published GISST algorithm was not used in this case because a cumulative vulnerability score for the entire highway corridor was not useful. Instead, vulnerability within the corridor was evaluated by summing vulnerability scores within 1 km^2 areas. The Panel notes that application of the GISST to evaluate the proposed highway system significantly limited the size of the appropriate geographic unit in the analysis. The comparison of possible highway routes, rather than determining a cumulative score for the highway corridor, was most valuable to planners, and hence the approach used by Region 6 is reasonable. Adding rather than multiplying indices (as EPA has done using the vulnerability criteria scores in the Interstate Highway 69 case) is acceptable if the indices are averaged, and as long as they represent only impacts or only vulnerabilities. The Panel recommends that the approach used in the Interstate Highway 69 case be refined and provided to GISST users as an optional algorithm. However, as discussed in the response to charge question 2.3 below, the GISST should only be used to conduct initial

evaluations of potential environmental impacts. Summary criteria in the GISST should not be used to make project decisions.

3.2 Question 1.2. Appendix A of the GISST User's Manual identifies the impact and vulnerability criteria that are used in the GISST to evaluate environmental impact and vulnerability. A subset of these criteria[3] is frequently used by EPA Region 6 to conduct GISST evaluations. Are the criteria in this subset reasonable and appropriate for use in GISST evaluations of the potential degree of vulnerability of a project area and the potential degree of impact produced by a proposed project? Please provide similar comments for the other criteria in Appendix A. Are there additional categories of criteria that should be developed for use in GISST evaluations?

The Panel finds that, given limited alternatives and available resources, EPA Region 6 has developed the GISST thoughtfully and implemented it with good balance. Comments concerning the use of the criteria and recommendations to improve the criteria are provided below. The Panel emphasizes that these comments do not imply that the GISST is inadequate for use in initial evaluations to detect potential environmental impacts. However, with the information currently available in the GISST User's Manual, it is difficult to answer the question of whether the criteria are "reasonable and appropriate." The Panel notes that several elements make GISST evaluations different from other GIS evaluations. Unique GISST elements include: 1) the criteria scoring process, 2) criteria subset selection, and 3) the process that highlights important drivers of concern for further analysis. These elements make the GISST an objective, spatially explicit tool for conducting initial, broad-stroke evaluations in a timely fashion. The Panel notes, however, that better definition and description of these processes is needed in the GISST documentation.

Many criteria are described in the long appendices of the GISST User's Manual. Each criterion is described in a one-page summary. The Panel understands that including detailed descriptions of the criteria in the User's Manual would make the document very long. However, the User's Manual currently contains insufficient information to provide a complete understanding of the basis for screening-level actions or decisions. The Panel suggests that more detailed explanations of the criteria and examples illustrating their use could be placed on a GISST website. Interested persons could obtain information from the website to gain a better understanding of a decision or discussion that was based on a GISST evaluation. As additional information for such a website is developed it will be important to keep the perspective of vested or interested parties in mind.

As the Panel deliberated on the charge questions it became apparent that decision analytic and statistical support for development of the GISST criteria and scoring system had been

[3] The subset of criteria most frequently used in GISST evaluations includes: Stream Density (surface water quantity), Population Density, Minority (environmental justice), Economic (environmental justice), Agricultural Lands, Density of Managed Lands, Hazardous Waste (Other Industries or Pollution Sources), Impaired Stream Segments (Clean Water Act 303(d) Segments), Wetlands, Floodplain, Ozone Nonattainment, Texas Ecological Assessment Protocol (TEAP) Diversity, TEAP Rarity, TEAP Sustainability, TEAP Composite, Wildlife Habitat, Federally-listed Species, and State-listed Species, and Ecologically Significant Stream Segments. The TEAP criteria were derived using a tool developed by EPA Region 5, the Critical Ecosystem Assessment Model (CREAM). The SAB has reviewed the CREAM. The SAB report on the CREAM is available at: http://www.epa.gov/sab/pdf/cream_sab-05-011.pdf

unavailable to EPA staff. The Panel strongly recommends that decision analytic and statistical support be provided to the GISST developers to revise and improve the criteria. Many questions and issues raised by the Panel could be remedied through discussions with statisticians. For example, a statistical examination of the data layers underlying the criteria could identify redundant measures and increase the soundness of rating scales.

Appropriateness of "most frequently used" GISST Criteria

The 19 "most frequently used" GISST criteria listed in the footnote to charge question 1.2 were used in the Interstate Highway 69 corridor evaluation summarized in the GISST User's Manual. These criteria appear to have been selected on the basis of some requirements for the federal highway impact assessment process and seem to be appropriate for use in this case, but the GISST documentation does not describe how or why these criteria were chosen for the case. There is an extremely heavy emphasis on vulnerability scores within this group of criteria and it is not clear how criteria that appear to cross the vulnerability and impact categories are used. Of the 19 frequently used criteria, 12 are indices of vulnerability, 3 are indices of impact, and 4 are listed as indices of both vulnerability and impact. Sorting these criteria into GISST groups that define the type of impact or vulnerability shows the following distribution among groups: 11 are ecology, 3 are water quality, 3 are socioeconomic, 1 is toxicity, and 1 is air quality.

The Panel notes that different subsets of criteria will probably be selected for use based on the context of the screening assessment to be conducted. The Panel recommends that, instead of focusing on the appropriateness of a particular group of criteria, it would be better to define a general process whereby groups of criteria could be selected for use in various kinds of evaluations. Different impact templates could be created and used as guides for selecting criteria to conduct different screening activities. Templates might be developed to describe the impacts from energy facilities, transportation projects and CAFOs. Each of these general types of projects would share a set of predictable impacts and the key criteria relevant to evaluation of those impacts could be identified.

Overall, the Panel finds that the criteria need to be better described in the GISST User's Manual. Information describing the criteria databases, references, definitions, limitations, and uncertainties is not clearly presented in a form that can be easily understood by general users. An exception is the Texas Ecological Assessment Protocol (TEAP) criteria group. Within the TEAP group, justification and definition of how the criteria are used is provided. Criteria definitions are especially important for GISST users who may be asked to define new specific groupings of criteria for future analyses. In the appendices of the User's Manual, the criteria are broken into groups that are defined according to types of impacts (e.g., water quality, toxicity). Criteria in these groups are then broken into vulnerability and impact categories. The criteria are presented in a somewhat arbitrary order within the impact groups.

Appropriateness of other GISST Criteria

In addition to the 19 frequently used GISST criteria, 50 other criteria are described in the GISST User's Manual criteria appendices. The Panel finds that these other criteria appear to be reasonable for use in some cases. However, the context of an assessment will drive a determination of whether the selected criteria are reasonable and appropriate. Without a better rationale for use of the criteria, and a more detailed description of how they may be used, it is

possible to provide only an intuitive response concerning the appropriateness of the criteria. Specific comments on some of the individual criteria are provided in Appendix A of this report.

The EPA Region 6 presentations to the Panel have illustrated an important feature of the GISST that highlights the usefulness of the criteria. This feature is the power of the GISST for use in conducting "scoping" studies to identify potential issues of concern. Such large scoping studies requiring use of all the criteria are likely to be very informative to applicants seeking approval of projects.

Comments on the Vulnerability and Impact Criteria Categories in the GISST

The Panel finds that the term "criterion" used in the GISST documentation might not be ideal because it often implies a threshold. The Panel suggests that "attribute" or "factor" would be better terms to use in this context. Also, the uses of the terms "vulnerability" and "impact" in the GISST documentation are questionable. As discussed below, different kinds of vulnerability and impacts are considered in the GISST (e.g., environmental justice and ecology). The Panel suggests that it would make sense to separately track broad categories of criteria (e.g., water quality, socioeconomic) with clearly stated objectives for conditions that are desirable. The Panel recognizes that it is important to balance competing thresholds in a NEPA assessment, but those thresholds need to be set out separately in order to make the best judgments.

A key issue to be addressed is the way groups of criteria are utilized in the GISST. The basic GISST algorithm is heavily dependent on the potential area affected by a project. As noted above, the number of criteria lumped into vulnerability or impact areas can affect the ability of important criteria to drive the assessment outcome. For example, the CAFO case study in the GISST User's Manual is driven by the areal extent of potential impact and not by vulnerability scores or impact scores. In addition, the way composite GISST scores are broken into categories seems to be arbitrary or not well described (e.g., the CAFO case study presents information on the composite score in a non-linear fashion with no justification or rationale). Many of the same comments can be made relative to the Interstate 69 Highway study scaling issues, composite scores, and the use of composite or group data.

Evaluating the degree of vulnerability is compromised to some extent in the GISST because the criteria aim to identify fundamentally different types of vulnerability in a single measure. In a very general sense, there seem to be three kinds of vulnerable systems with which EPA is concerned: 1) stressed, heavily disturbed systems that have already sustained relatively high cumulative impacts and have limited (or no) capacity to absorb additional stressors ("type 1 vulnerability")[4], 2) relatively unaltered or "pristine" areas that may be regionally important because they represent the few remaining systems or places with high ecological value ("type 2

[4] The following GISSTcriteria are examples of measures that primarily reflect "type 1" vulnerability because higher scores are associated with more contaminated, altered, fragmented, and disturbed environments (note that for these criteria, "type 2" vulnerable areas would receive low Dv scores): water releases, ground water quality, channelization, agricultural lands, habitat fragmentation, road density, population density, total population, other industries, pollution sources, or protected lands, hazardous waste, impaired waters, and ozone nonattainment areas.

vulnerability")[5], and 3) areas that have unique characteristics (hydrologic, human, etc.) that would make these areas or population in them particularly vulnerable to contamination ("type 3 vulnerability")[6]. The first two types of vulnerability are especially problematic because they are on the opposite ends of the spectrum in the GISST scoring process. For example, the vulnerability measures for an agricultural landscape with fragmented habitat, high population density, and many roads would be higher for many of the criteria than for an ecologically important forested landscape with large blocks of forest-interior habitat, little anthropogenic disturbance, and high water quality. Several alternative approaches might be considered to address this problem. One alternative is to calculate three different vulnerability indices. Another alternative is to separately consider the ecological value of an area (i.e., what is present that we want to protect?), vulnerability to negative impacts from project (i.e., what is the likelihood that the system will be impacted by the project?), and magnitude of change in system (i.e., how much and in what way is the system likely to be altered?).

General Comments on the GISST Criteria

The Panel provides the following general comments on the criteria. Additional comments on specific GISST criteria are included in Appendix A of this report.

There is a need to better delineate how to use the groups of criteria. It is presently unclear how one decides to include or exclude particular criteria in the analysis. The GISST documentation should include more quantitative analyses to support the basis for choosing criteria and their weights. Substantial information concerning selection of indicators is available in the literature (e.g., Cairns et al, 1993; U.S. EPA Science Advisory Board, 2002; National Research Council, 2000). Cairns et al. (1993) have reviewed how different types of indicators are used in decision-making. This paper may be particularly relevant when considering how socioeconomic criteria should be used in the GISST. As discussed below, the criteria weights and scoring system could be further developed through statistical analysis of information provided through an expert opinion process.

- The individual GISST criteria each make intuitive sense but, as noted above, they can have competing purposes and counteract one another. They can also have different thresholds (i.e., different values) and may be combined in illogical ways (e.g., combining noise and odor criteria scores with scores for use of energy efficient appliances). There is a need to define criteria categories in a consistent manner whereby combinations or groupings make sense in an overall strategy to achieve assessment objectives (e.g., assessing current status, different kinds of vulnerabilities, future impacts).

[5] The following GISST criteria are examples of measures that reflect "type 2" vulnerability because higher scores are assigned to environments with greater ecological "value" based on measures of diversity, rarity, and sustainability (note that for these criteria, type 1 vulnerable areas would receive low Dv scores): wildlife habitat, TEAP diversity, TEAP rarity, TEAP sustainability, and TEAP composite.

[6] The following GISST criteria reflect type 3 vulnerability because higher scores are assigned to areas with unique or defining features likely to make the area or human population residing in the area particularly vulnerable to negative impacts from the project: wetlands, floodplains, stream density, surface water, surface water quantity, distance to surface water, ground water probability, average stream flow, sole source aquifer, aquifer/geology rating, individual well water, septic tank and cesspool use, soil permeability, high school education, colonias, educational achievement ranking, economic, minority, age, children, older population, pregnancy, houses lacking plumbing, ability to speak English, foreign born, telephone communication, and linguistic isolation.

- As EPA continues to develop the GISST criteria, the Agency should consider incorporating measures of runoff potential and nonpoint source pollution.

- A fundamental concern of the Panel is the relevance of the GISST criteria to specific projects. EPA has stated that some of the criteria were incorporated into the GISST because data were readily available, not necessarily because the criteria were relevant to a problem or management objective. As suggested above, it might be possible to improve the GISST by developing alternative "impact templates" that identify the key criteria relevant to particular types of projects. Impact templates may reflect different models that are applicable to specific types of projects. Such models might be developed by conducting expert assessments of impacts and vulnerabilities for various types of projects and by using statistical methods to evaluate the explanatory variables (i.e., criteria).

- Although a large number of spatial overlays representing different criteria can be viewed in the GISST, it is important that analysts use only the most significant layers in an assessment. The significance of truly critical factors can be diminished if so many are equally considered, and it may become difficult to determine how a particular result was obtained. In its application of the GISST, EPA should keep in mind the Principle of Parsimony and not increase, beyond what is necessary, the number of entities required to explain anything. EPA should select and use criteria that are significant and critical to decisions and processes. The selection of criteria that are "significant" can be validated by conducting a statistical analysis of data on the outcome variable of interest (i.e., vulnerability or impact) for a random sample of plots for which the explanatory GIS data and values of the outcome variable are available.

- As discussed above, the Panel is also concerned that EPA has not clearly described differences between criteria specified to determine the D_v and/or the D_i. Some GISST criteria reflect both D_v and D_i and other criteria reflect only one of these indices. There are several cases where it is unclear whether impact or vulnerability is the real focus of a particular criterion. For example, the toxicity-weighted water releases criterion on page A-62 of the GISST User's Manual is identified as an impact criterion, but it is not clear why it is not also a vulnerability criterion.

- As stated above, the Panel finds that the GISST User's Manual, particularly Appendix A, would greatly benefit from additional detail regarding the rationale for, definition of, and assumptions underlying specific criteria. Definitions are needed in some of the descriptions of individual criteria in the User's Manual. For example, the criteria descriptions of ecologically-significant stream segments and density of managed land make references to conditions within an "area" but fail to specify the area. In some cases, the rationale for scoring is not obvious to the reader/user.

3.3 Question 1.3. The GISST uses data sets (in Appendix A) with different coverages generated for different purposes (e.g., point sampling of water quality, census data, and land cover data gathered by satellite). Is the GISST 1 – 5 scoring scale on these coverages and datasets reasonable for developing an initial assessment of the potential cumulative impacts of proposed projects?

The Panel discussed the use of different criteria and coverages in the GISST 1-5 scoring system. An advantage of the GISST 1-5 scoring system is that it allows (or forces) the evaluators to simplify the combination of diverse criteria and show sharp contrasts in the assessment of potential impacts or comparisons among project alternatives. A limitation of the scoring system is that assessors lose the ability to see what specific information the underlying data provide for a particular assessment. In EPA's description of the scoring system, it is not clearly stated that some criteria scores are based on graded responses (on a true 1-5 scale), some criteria scores are based on intermediate scales (1, 3, 5), and some are based on binary scales (1, 5). The type of scale used to score each of the criteria is not clearly defined so it is not possible to evaluate the impact of the scoring system on the final outcome of an assessment.

The Panel questions whether the GISST criteria can be generically scored without considering the type of project being evaluated. For instance, the influence of rainfall or stream density on vulnerability may be quite different in a coal-fired electrical generation plant evaluation and a swine CAFO project evaluation. As discussed above, EPA may need to develop a set of scoring templates that incorporate criteria associated with the general types of projects that are evaluated. Each template would better incorporate scientific understanding of the processes that the criteria represent for a specific type of project. The Panel has identified the following specific limitations and concerns regarding the GISST 1-5 scoring system.

Scoring system limitations and issues of concern

- As stated above, the Panel is concerned that some of the criteria in the GISST User's Manual may not be relevant to specific projects. Similarly, it is not clear that the 1-5 scores assigned to criteria are always applicable for any project evaluated. The appropriateness of some of these criteria scores may be "project-dependent."

- The Panel questions whether all of the criteria should always be scored with a maximum value of 5. It seems that a score of 5 assigned to any criterion should "raise a red flag" and be indicative of a problem. A score of 5 may not be appropriate for criteria that are less important in an evaluation. This is not to suggest that the potential range change, rather it means that the value of '5' may not always be observed.

- As stated above, the 1-5 GISST criteria scores are summed and averaged. Averaging helps bring the scores into a common range when different numbers of criteria are used, but it also ultimately reduces the dynamic range of the results. In some sense this reduces the level of information, but the values could be rescaled to accommodate this concern. It should be noted, however, that changing scales may add uncertainty to the analysis. It is helpful to preserve information about the distribution of component values that make up the averages. The GISST 1-5 scores appear to serve the purpose of numerical bins rather than simple rank-order ordinal measurements. The only permissible arithmetic operation using ordinal values is transitive relationships. Addition, subtraction, multiplication, and division using ordinal numbers do not produce valid or reliable results. Consequently, use of ordinal values in the scoring process is inherently invalid. The interval scale scoring method is reasonable for use in scoping level assessments to identify environmental concerns that are likely to require detailed analyses in subsequent

environmental impact assessments. For this use, the scores are reasonable for detecting potential individual concerns where no mathematical operations are required for aggregating scores. Scoring methods similar to the one used in the GISST have been in widespread use in suitability analyses since the 1960's (McHarg, 1969) for quantifying the relative desirability (or vulnerability) of sites for project alternatives. However, such an application of the GISST scoring system would require the use of a ratio measurement scale rather than an interval measurement scale.

- Most of the scaling of GISST criteria is based on physical values with true ratio quantities. Therefore, binning these values to the GISST 1-5 scale may be acceptable. However, the values seem to have been binned or scaled using different techniques or functions: linear, interval, stepped, logged, and natural breaks. An assumption in the GISST is that the scores of different criteria are equivalent, but this may not be the case. If the values are scaled using different techniques or functions, it is invalid to equate the scales of different summed values. There is no easy way to address this problem unless it can be demonstrated that the values of scaled variables are equivalent (e.g., a value of 2 for one scaled variable is equivalent to a value of 2 for another).

- There is probably a large difference in the scales of spatial data layers used in the GISST. The combination of spatial data that are highly variable in scale may create a mixture of incompatible spatial frequencies since phenomena are scale-dependent.

- A scoring system with a scale of 0-5 might be more appropriate than the 1-5 scoring system because some features represented by the scores may be associated with no vulnerability or impact. These features should receive a score of zero.

- Scoring the 19 criteria listed in charge question 1.2 on the GISST 1-5 scale provides the potential for a cumulative score of 95. Any one criterion is therefore insignificant because it can only influence the total score by 5 points. As noted above, it is important to identify the criteria that are most critical for an evaluation and make sure they do not "get lost" in the scoring process. Various methods for doing this, such as weighting, have been discussed.

- The weighting of GISST criteria has been discussed above. It is the use of GISST, and not particular numbers calculated, that most contributes to the decision-making process. Therefore, the weighting of individual criteria must be performed correctly. Typically this is accomplished by developing specific weighting factors (e.g., multiplying by 1, 2, 3, etc.). There are inherent problems associated with this approach, and they must be considered. For instance, in the GISST vulnerability associated with a particular criterion is scored on a scale of 1-5 where 1 represents the lowest vulnerability and 5 the highest. If scores on this scale are multiplied by a weight of 2 the bad score of 5 becomes two times worse, but the better score of 1 also becomes worse by a factor of two (since higher values represent more vulnerability). This is problematic because assigning a weight of 2 to the GISST score of 1 should make it two times better. Normalizing the weights on a scale of 0-1 will alleviate this problem.

- The GISST does not account for the possible interaction of data layers used in the tool. Some layers may be correlated (e.g., soils and geology) or have compound or inverse relationships (e.g., wetlands are typically not found at high elevations). The decision of whether to include a data layer in an evaluation process should be based on the environmental context of the data layer and its statistical relationship to other data layers. Furthermore, without differential weighting the GISST approach runs the risk of improper double counting of information. Methods for constructing relational trees (diagrams) may be helpful in evaluating the interaction of data layers.

- When landscape features are rescaled or aggregated, the Modifiable Areal Unit Problem (Openshaw, 1984) is introduced. EPA should be aware that values representing variation in an attribute can be significantly modified by cell size, resolution, and aggregation based on boundaries, categorization, or grouping. The following potential effects should be considered and discussed when the GISST data are grouped or categorized in areal units to assign scores to the criteria:

    - Scale effect (raster resolution). This effect is a variation in numerical results caused by grouping data in different zone sizes.
    - Aggregation or zonation (raster resampling). The effect is a variation in numerical results caused by grouping zones, methods used, or differences in zonal boundaries.

Suggestions to improve the scoring system

The Panel provides the following suggestions to improve the GISST scoring system.

- As suggested above, rather than combining the scores of the all GISST criteria, it may be more informative to group the scores on the basis of similar characteristics such as air quality or water quality. The six broad categories of ecological indicators identified in the SAB Framework for Assessing and Reporting on Ecological Condition (U.S. EPA Science Advisory Board, 2002) could be separately and independently scored (i.e., landscape, biotic, ecological processes, chemical & physical, and hydrology & geomorphology). Perhaps the vulnerability (D_v) and impact (D_i) criteria groups could be scored independently and separately, not summed or averaged with each other. This approach would give the decision-maker more information and multiple scores instead of a single score based on different criteria. It may also be useful to highlight the scores of "critical" criteria.

- The process of summing the scores of the GISST criteria is similar to counting the number of criteria that have been assigned various scores. Rather than using summed scores, it may be more informative to combine the criteria using such score counts and generate maps showing the number of features that have been assigned various scores. This approach preserves information about the distributions of scores across the different criteria that make up a group.

- Critical features could be mapped using criteria functions with Boolean operators. In this

sense, the user defines values that identify significant vulnerability or impact for each of the criteria and they are combined on this basis as a cumulative map.

- Where no data are available, GISST assigns criteria values differently among the criteria. The "no-data" areas should be treated consistently, perhaps flagging such sites as "potentially vulnerable".

- With respect to the three vulnerability types mentioned previously, alternate scoring systems could be applied to the same criterion. For example, the scoring system for the criterion "habitat fragmentation" currently assigns higher scores to more fragmented landscapes. This is appropriate for identifying systems that have already sustained high levels of impact, but the criterion could also identify areas with high ecological value (type 2 vulnerability) if the scoring system were reversed (i.e., areas that are not fragmented receive high scores).

Statistical issues concerning the GISST scoring system

Several statistical issues emerge relative to the use of the GISST criteria and scoring system.

- Many of the criteria seem to overlap and the potentially important consequences of this are unclear. A formal analysis of these consequences is needed. D_v and D_i are derived by multiplying or summing the scores of "mixtures" of criteria that are often selected based on the availability of underlying data. The criteria are quite variable in the content of underlying data, validity, and utility. Some of the criteria are quite specific, and some are quite broad.

- The combining of criteria scores may violate assumptions of independence of the data or be influenced by data correlations that should be investigated.

- The criteria scoring system could be significantly improved by involving more statistical, expert elicitation, and experimental design expertise. Natural breaks in the information were used to develop the GISST scoring system based on a software decision algorithm. That approach might not be as effective or desirable as using an expert opinion-based ranking system. This is because an expert opinion process could potentially provide more insight and meaningful measures of uncertainty. The Panel notes that it is important to carefully consider how such expert opinion would be gathered. EPA may want to consider asking each expert to independently develop judgments and scores. A central tendency could then be used instead of gathering experts together to obtain a collective assessment. If independent expert opinions were collected, some measure of the uncertainty in the scores could be generated and potentially used to discard or retain various criteria, or to stimulate further attention to areas with inadequate information. Experts can develop judgments about the impact or vulnerability associated with actual plots of land in a random sample of plots for which the GIS variables are also available. These independent ratings could be used as dependent variables to calibrate statistically derived criteria weights. The Panel recommends that EPA Region 6 staff explore the use of expert opinion elicitation methods, pattern recognition techniques, and visualization

tools to revise the criteria scoring process.

- It became apparent during EPA's presentations to the Panel that the Agency had used implicit rules and informal criteria weightings in their GISST analyses. However, the rules for decisions involving appropriate "show stoppers" or moving the decision process out of a straightforward GISST scoring system were unclear. The Panel notes that the lack of clear, formal rules or criterion weightings could produce stakeholder or EPA confusion in the decision process. The Panel therefore recommends that explicit decision rules and criteria weightings be clearly defined.

3.4 Question 2.1. EPA intends to use the GISST in the NEPA process as an initial screening tool to prioritize potential single, direct, and cumulative environmental impacts of projects for more detailed analyses. Please comment on the strengths and limitations of the GISST as it applies to this purpose.

The Panel finds that, with the following caveats and recommendations, the EPA Region 6 Geographic Information System and *some* of the GISST criteria can be used in the National Environmental Protection Act (NEPA) process to assist the Agency in reviewing and scoping environmental assessments (EAs) and environmental impact statements (EISs).

- The GISST should be used solely as an initial environmental impact identification tool. In this regard, the GISST should only be used as a tool to detect potential impacts commonly associated with certain types of projects.

- Sets of core criteria should be identified (from the list of frequently used criteria provided in charge question 1.2) for use in evaluating certain industrial activities or projects. For example, it is likely that sets of core criteria could be identified to evaluate projects such as CAFOs, highway construction, coal-fired power plants, and other kinds of projects. These industry-specific lists should be made publicly available.

- Decisions based upon single, combined criteria scores should be avoided for all GISST applications. Summing the scores of specific groups or categories of criteria may be appropriate (e.g., toxicity, water, environmental justice/socioeconomics), but scores should not be summed across these categories. For example, it would be inappropriate to combine "Noise and Odor Thresholds" with "Energy Efficient Appliances." Depending on the project, water quality impacts are likely to be on a different scale of importance from odor/noise or energy efficient appliances. These groups of criteria scores should not be combined, or at least not without establishing some weighting system (other than the system of equal weights that is currently applied). The GISST summary algorithm should never be used in and of itself to make a screening decision.

- Cumulative scores, if used, should be very clearly defined and communicated to NEPA applicants. The GISST should be made available to the public so that applicants can evaluate their data in advance of permitting activities. EPA's NEPAssist program may be the place to develop this option.

- While the GISST may be useful for a preliminary evaluation of environmental impacts, it is not adequate for prioritizing impacts unless the following elements can be changed or addressed:

    1. Weighting of the criteria. The GISST must in some way reflect how the scores represent the relative importance of each impact (i.e., the values of the decision-maker). Since the criteria, and hence the criterion scores, are equally weighted (each is multiplied by 1) it is all but assured that such equal weights do not reflect the relative importance that decision-makers or stakeholders would accord them, and they do not reflect society's values. Since prioritization reflects the values of an individual or group of individuals, without deliberate weighting of the criteria the GISST should not be used for prioritization. It is impossible to make decisions about "impacts" without linking them to goals and objectives

    2. Stakeholder concerns. GISST criteria must reflect the concerns of stakeholders or decision-makers for the specific problem being addressed. The 19 frequently used criteria in the GISST do not necessarily reflect the problems or projects to which they will be applied and the concerns (values) that people may have about these projects. Some of the 19 criteria may not be relevant at all. Since data sets are selected for an evaluation before the problem (objectives and concerns) is defined, priorities identified would be highly suspect.

    3. Eco/physiographic regions. A scoring system must be developed to reflect the impacts and vulnerability in specific homogeneous ecoregions and physiographic regions where the scores will be applied. The scoring system cannot be applied universally across EPA Region 6. For example, the scale developed for ranking stream density in the arid Texas panhandle or Oklahoma would not be appropriate for use in humid southeastern Texas or Louisiana (Land Evaluation, 1968). Prioritizing impacts on the basis of those value scores would likely be inaccurate and misleading. In this regard a classification system different from the hydrologic units (HUCs) could be developed or adapted. An approach that might be considered for some uses is the ecoregion system (Omernik, 1995).

    4. Inaccurate representation of cumulative impacts. Cumulative impacts are represented in the GISST by averaged criterion scores (for vulnerability and impacts separately), summed criterion scores (for the Interstate Highway 69 case), or the scaled product of average criterion scores (i.e., the total "CRIA" score yielded by the GISST formula). The compensatory trade-off nature of these computations, the lack of weighting, and the use of inadmissible operations on ordinal or interval level scores (discussed above) will mean that cumulative impacts will be inaccurately measured and priorities may be misguided.

    5. Inaccuracies of modeled impacts arising from the use of imprecise data scales. EPA should be aware of inaccuracies that can result from using imprecise scales of data. The scale of resolution of the information and analysis should be appropriate to the scale at which the decision is being considered.

6. Selection and documentation of map classes. Because different map classes will show different patterns and hence indicate different impact priority areas, map classes selected should be transparent and well documented. Through graphical pattern display, geographic information systems have the potential to effectively and efficiently divulge differential spatial impacts that can highlight high priority areas of environmental impact. In the GISST such patterns depend on the class intervals used in choropleth maps. The GISST produces and generates those intervals automatically by a data-determined algorithm (Jenks & Coulson, 1963) (rather than by using people's values and concerns) that users are not likely to understand. While this provides a good tool for exploratory impact identification, users may not employ it in this way and no guidance is provided to interpret the GISST outputs.

7. Consideration of spatial interdependence of processes. The manner in which GIS is used in the GISST fails to exploit its capabilities as a spatial analysis tool, limiting its use mainly to spatial display or mapping. Spatial interdependence and interaction of environmental and socioeconomic processes and consequent effects appear to be ignored in the GISST. Computations in the GISST are done on a cell-by-cell (or polygon-by-polygon) basis without considering the effects of the same or similar processes operating in nearby cells. Thus the determination of cumulative impacts due to interaction effects is likely to be missed. The Panel notes that truly accounting for spatial interaction of environmental processes is equivalent to modeling and it would require involving additional experts. However, this may be beyond the scope the GISST, which is not intended to be a predictive modeling tool.

3.5 Question 2.2. EPA also intends to use the GISST in the NEPA process to evaluate environmental impacts of project alternatives to help inform decision-making. Please comment on the usefulness of the GISST as a tool for this use.

As stated above, the Panel finds that it is reasonable and appropriate to use the GISST as a tool for broad-stroke preliminary assessments of large complex projects. In these cases, individual criteria, or suites of criteria, could be evaluated to detect potential environmental impacts that might be associated with project alternatives. Such preliminary evaluations can help inform decision-making. However, proper evaluation of project alternatives requires comparison of the projects with respect to different dimensions of importance (the criteria) and determination of relative advantages and disadvantages of each project. Limitations of the GISST discussed above impede or prohibit such a comparison. The GISST is further limited for use in evaluating the environmental impacts of project alternatives because the criteria selected for use, and scaling of the underlying data, must be specific to the ecoregion in which they are applied. For example, in the Interstate Highway 69 case the stream criteria scaling that was applied to the corridor in Brownsville, Texas, is not an appropriate scaling for application in the Tyler, Texas, corridor. These scaling issues limit the usefulness of the GISST in large scale assessments for decision-making such as the Interstate Highway 69 corridor study.

To make the GISST useful for evaluating the environmental impacts of project alternatives to inform decision-making the tool should be tailored to address the following questions and issues.

Question 1: Does a project have potential environmental impacts?

Issues of potential concern to decision-makers:

- Summary scores that do not include weighting would potentially mask environmental impacts.

- All criteria that are relevant to the stakeholders and decisions makers must be included in the analysis.

- Irrelevant criteria must be excluded.

- Relevant and meaningful thresholds or impact levels should be considered (e.g., stream criteria applied to Brownsville vs. Tyler, Texas).

- Spatial dependence of the cells in the GISST must be explicitly considered.

Question 2: Is it easy to see why a project has an impact?

Issues of potential concern to decision-makers:

- The current GISST summing algorithm cannot identify the impacts. Decision-makers need to examine individual criteria scores and their associated maps to identify impacts.

- The vulnerability and impact criteria can be confused.

Question 3: Is it easy to see the advantages/disadvantages of alternative projects?

Issues of potential concern to decision-makers:

- It might be important to map two project alternatives simultaneously. Although GIS clearly has such capability, it is not evident that the GISST allows one to create "difference maps" of the criterion scores pertaining to two different alternatives. Likewise, with three or more alternatives, it is not clear that the GISST could make a synthesis map showing which project had extreme criterion scores for any particular geographic region (e.g., show which project would have the worst water quality in each grid cell). There is a need for improved graphical presentation.

- The sum of GISST averages does not provide complete information about the pros and cons of alternative projects. Two projects may have equal cumulative scores but it is important to understand how low scores on individual criteria may affect the cumulative score. The GISST does not enable one to conduct sensitivity analyses easily. Since differential weightings are not allowed, one cannot investigate how different value tradeoffs would lead to preferences for different projects. Likewise, the 1-5 value scores assigned to the raw score intervals cannot be changed easily, so differences in expert judgment regarding vulnerability or impact cannot be investigated.

Question 4: Can the method be helpful in designing a new alternative?

Issues of potential concern to decision-makers:

- The method will not be helpful in designing a new alternative unless the criteria and objectives are defined *a priori*.

3.6 Question 3.1. Please provide recommendations on steps that can be taken to enhance the usability of the GISST User's manual and documentation.

The Panel finds that the GISST User's Manual contains a good introduction to the GISST and the case example in the document helps the reader understand how the tool works. However, the GISST User's Manual does not contain step-by-step instructions on how to use the tool. The User's Manual provides documentation of the GISST, but it is not a user's manual. It appears to vacillate between providing specific details about how the GISST has been deployed and information about the general usefulness and applicability of the tool. The Panel finds that it would be very useful to develop a true user's manual for the GISST and offers the following suggestions and recommendations to improve the existing document. Additional specific comments on the User's Manual are provided in Appendix B of this report.

- Identify GISST users. It is not clear who a GISST user might be. The User's Manual does not provide adequate instructions on how to operate the tool nor interpret the outputs. The User's Manual provides a moderate amount of background description about the general approach, what it does, and the specific criteria that may be used. If the EPA team in Region 6 is the only user of the GISST, the existing User's Manual may provide them with enough information. However, if others will be using the tool, additional information must be provided. It is important to clearly identify GISST users and organize the document appropriately.

- Describe the need for the GISST, explicitly identifying the types of decisions EPA managers make in the NEPA process and how the GISST assists them in making those decisions. The User's Manual should provide background information describing the pressing need for this kind of tool to support NEPA assessments. The Manual should state that development of the GISST is a first step in meeting this need. Chapter 2 of the existing User's Manual contains criticism of approaches like GISST. For example page 13 provides criticism of the simple type of data integration applied in the GISST (Suter, 1993). Chapter 2 also lists some properties of GIS assessment tools (Leibowitz et al., 2000). However, the Manual does not respond to the criticisms nor identify how GISST was designed to achieve the desirable properties. In other words, it cites literature without interpreting it in the GISST context. The document needs to respond to these issues. The User's Manual does contain a very helpful comparison of the GISST criteria to indicators in the SAB framework for reporting ecosystem condition (U.S. EPA Science Advisory Board, 2002).

- Provide the underlying conceptual model. The User's Manual would be vastly improved if the authors began with a diagram/description of the conceptual model underlying the GISST, such as dose-response or EPA's risk assessment model. The criteria could then be related to the conceptual model. It is unclear why some GISST criteria represent

vulnerability and some represent impacts.

- Provide the basis and process for considering and selecting criteria. The basis and process for considering and selecting criteria for evaluation of different kinds of projects should be provided. EPA should consider organizing the criteria hierarchically into themes rather than a large number of unrelated criteria that make interpretation and synthesis difficult. The rationale for criteria scaling and binning should also be provided. For example, the criteria in Appendix A of the User's Manual could be organized according to vulnerability and impact, and then by criteria groups (e.g., water quality, ecological, air quality, socioeconomic, toxicity, etc.)

- Provide definitions. As noted above, vulnerability and impact are never defined in the User's Manual. For many of the criteria it is not clear why they are classified as impact or vulnerability measures. In fact, some criteria are listed as both types. Definitions of acronyms should also be provided in a glossary.

- Provide an index to the User's Manual.

- Clearly present the mathematical expressions. As noted above there are errors and ambiguities in the mathematical expressions describing the GISST algorithm.

- Define the mathematical boundaries of output parameters and provide guidance on interpretation of the range of expressed parameters.

- Provide suggested approaches for integration of spatial data. As noted above, the User's Manual provides criticism of existing approaches to integration of spatial data. It would be helpful to suggest useful approaches that might be applied to integrate the data.

- Name the criteria in terms of impact or vulnerability factors. The criteria are named in terms of spatial data rather than as impacts or vulnerability factors. As noted previously, the spatial data themselves are not impacts or vulnerability factors. This makes it difficult to understand the purpose of each criterion and whether a larger value in the raw data is good or bad. The description of each criterion should start with the relationship of the criterion to impact or vulnerability and not just the measurement units of the data variable. For instance, a criterion might be named "level of potential infiltration from rainfall" rather than simply "rainfall." Sometimes such a description is provided in the Definitions/Assumptions, but not always. As an example, the sole source aquifer criterion description on page A-13 does not explain the basic issue, which is that "a sole source aquifer makes the communities it serves vulnerable to project impacts." This explanation should be prominent at the beginning of the section. The section could then describe how spatial data are interpreted or analyzed to measure the criterion and provide the assumptions and limitations of the data or the mismatch between the data and the true meaning of the criterion (i.e., if the data are an indirect estimate of the criterion value).

- Provide examples of different applications of the GISST and relate them to the types of decisions mentioned above. It would be useful to provide at least one example for each

type of decision. Such examples could be organized around the uses of the GISST (e.g., scoping, designing alternatives, reviewing environmental impact statements, etc.) This would inform users of the appropriate use of the tool for each level of review. The GISST methods, criteria, and algorithm are intermixed in the User's Manual without identification of when and where their use is appropriate. The CAFO swine case study presented in the User's Manual is not very informative. The conclusions about specific impacts, monitoring, and denial of the application are not clearly supported by the brief description and maps, nor does the case study describe how to use GISST to make these determinations. The Interstate Highway 69 case study may not be the best one to use as an exemplar since it is stated on page 48 of the User's Manual that the methodology was changed for the analysis (i.e., the area and impact portions of the algorithm were not included, only vulnerability was considered, and scores were calculated for individual cells).

- Clarify the context of the NEPA Document Preparation and Review Section of the User's Manual. This section of the document was included after the conclusions. Although the section is listed separately in the table of contents, it is not clear what the section is supposed to illustrate. It does not appear to be a new case study nor does it appear to be part of the Interstate Highway 69 case study. The section needs an introduction with background material referring to a particular case study and showing how the information provided is different from material in previous sections. Table 3 in this section should be organized by vulnerability and impact criteria and criteria groups (see recommendation above concerning criteria names and descriptions) rather than providing a long list of criteria. The items included in the table are described as spatial data, not as criteria, and it is not clear how the data are interpreted (e.g., Is percent of population under 7 years of age a good or bad thing for a wildlife refuge? What does rainfall total mean?).

- Edit the final document. The final GISST User's Manual should be reviewed by a technical editor to correct such problems as inconsistent use of plurals (e.g., criteria vs. criterion), incomplete sentences, and typographical errors. The document should also be reviewed to ensure that technical words such as normalize, standardize, significant, and rank have been used correctly.

- Reorganize the GISST CD. The GISST CD should be reorganized to clearly indicate where the reader could locate the files containing the entire document and files containing various figures in the document. It is recommended that the entire document, including figures, be provided in one file for users who wish to print it. The document should also be placed on the GISST Internet website.

4. CONCLUSION

The Panel commends Region 6 for developing the GISST. Geographic Information System capabilities and data layers provide essential support for efficient, timely, and proactive NEPA evaluations and other Regional responsibilities. The benefits of compiling data layers for the GISST have undoubtedly extended to other applications within Region 6.

The Panel has identified a number of limitations of the methodological approach used in the GISST. These limitations restrict the usefulness of the GISST and must be considered in any application of the tool. It is reasonable and appropriate to use the scores of individual GISST criteria, or suites of criteria representing types of vulnerability, in preliminary evaluations to detect potential environmental impacts associated with certain types of projects. However, the aggregate GISST score should not be used in assessments for decision-making. The Panel has recommended improvements to make the GISST suitable for this use.

GIS-based assessment tools are needed to provide essential support for many EPA activities. Various EPA program offices and regions have developed screening tools similar to the GISST since GIS technology became widely available in the 1990s. Examples include the EPA Region 4 Southeastern Ecological Framework, the EPA Region 5 Critical Ecosystem Assessment Model, the EPA Office of Water's Index of Watershed Indicators, the EPA Office of Research and Development's Regional Vulnerability Assessment methods, and the NEPAssist web-based mapping tool developed by EPA's Office of Federal Activities. The compartmentalized development of GIS tools and data by EPA program offices and regions is inefficient, given the universal value of these tools. The Panel strongly urges EPA to undertake a concerted effort to develop a unifying framework for the development of assessment tools based on spatial information technologies.

5. REFERENCES

Cairns, J., P.V. McCormick, and B.R. Niederlehner. 1993. A proposed framework for developing indicators of ecosystem health. *Hydrobiologia* 236: 1-44.

Jenks, G.F., and M.R. Coulson. 1963. Class intervals for statistical maps. *International Yearbook of Cartography* 3:119-134.

Land Evaluation. 1968. Edited by G. A. Steward. MacMillan. Melbourne.

Liebowitz, S.G., B. Abbruzzese, P.R. Adamus, L.E. Hughes, and J.T. Irish. 1992. *A Synoptic Approach to Cumulative Impact Assessment: A Proposed Methodology*. EPA/600/R-92/167. Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

Leibowitz, S.G., C. Loehle, B-L. Li, and E.M. Preston. 2000. Modeling landscape functions and effects: a network approach. *Ecological Modelling* 132:77-94.

McHarg, I.L. 1969. *Design with Nature*. Natural History Press. Garden City, NY.

National Research Council. 2000. *Ecological Indicators for the Nation.* Committee to Evaluate Indicators for Monitoring Aquatic and Terrestrial Environments, Board on Environmental Studies and Toxicology [and] Water Science and Technology Board. National Academy Press, Washington, D.C.

Omernik, J.M. 1995. Ecoregions: A spatial framework for environmental management. Pages 49-69 in W.S. Davis and T.P. Simon, eds. *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*. Lewis Publishers, Boca Raton.

Openshaw, S. 1984. The modifiable areal unit problem. *Concepts and Techniques in Modern Geography* 38:41.

Suter, G.W. 1993. A critique of ecosystem health concepts and indices. *Environmental Toxicology and Chemistry* 12:1533-1539.

U.S. EPA Office of Science Policy. 2003. *Draft Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models*. U.S. EPA Office of Science Policy, Office of Research and Development, Washington, D.C. (http://www.epa.gov/ord/crem/library/CREM%20Guidance%20Draft%2012_03.pdf)

U.S. EPA Science Advisory Board. 1999. *An SAB Report: Review of the Office of Water's Index of Watershed Indicators.* EPA-SAB-EPEC-99-014. (http://www.epa.gov.sab/pdf/epec14.pdf)

U.S. EPA Science Advisory Board. 2002. *A Framework for Assessing and Reporting on Ecological Condition: An SAB Report*. Edited by T. F. Young, and S. Sanzone, EPA-SAB-EPEC-02-009. U.S. Environmental Protection Agency, Washington, D.C. (http://www.epa.gov/sab/pdf/epec02009.pdf)

U.S. EPA Science Advisory Board. 2005. *Review of the EPA Region 5 Critical Ecosystem Assessment Model (CrEAM*). EPA-SAB-05-011. U.S. Environmental Protection Agency, Washington, D.C. (http://www.epa.gov/sab/pdf/cream_sab-05-011.pdf)

Wiens, J. 1996. Oil, seabirds, and science: The effects of the Exxon Valdez oil spill. *Bioscience* 46:587-597.

Appendix A. Specific Comments on individual GISST Criteria

- Rainfall (page A-4). The rationale for increasing vulnerability score with increasing rainfall is not clear. Might some xeric areas actually be more vulnerable to projects and consequent changes in hydrology?

- Water releases (page A-5). TRI releases only consider releases onsite, not flows from upstream units.

- Surface water quantity (page A-6). The water quantity criterion is only based on length of streams, not on volume. This should be validated where data are known for actual volume. Otherwise this criterion may be a poor proxy.

- Distance to surface water (page A-7). EPA could use GIS flowpath algorithms to get distance instead of using straight-line distance. Again this criterion is a poor proxy.

- Ground water probability (page A-8). It is not clear what probability this criterion measures. Is this the proportion of cells where the water table is less than 6 feet from surface? It would be helpful to include an explanation of the rationale for the 10-acre scale and the 6-8 foot threshold in the definition of the criterion.

- Unified watershed assessment (page A-10). This criterion is defined exactly like the surface water criterion described on page A-2 but with different scoring. Are they correlated? Differences should be explained.

- Floodplain, and others where a zero value indicates no data available (page A-14). The use of zero values for missing data is acceptable if the only use is to flag the lack of data. It is inappropriate to actually use the zero value in the analysis since it unbalances the composite scores whether they are summed or averaged.

- TRI reported water releases (page A-19). TRI releases are listed as both an impact and vulnerability criterion. This appears to be a repeat of the water release criterion provided on page A-5.

- Soil permeability (page A-20). The definitions and assumptions are copied from the water table depth criterion and are not relevant for this criterion.

- Agricultural lands (page A-21). This measure is very coarse because it combines all agriculture types that may differ greatly in associated vulnerabilities.

- Agricultural land (page A-21). It is not clear how farmland can be both a vulnerability and impact criterion. Is it even useful to include this criterion without mapping prime farmlands? The U.S. Department of Agriculture and/or states have mapped prime farmlands and other classes for the U.S. It is not necessary to rely on National Land Cover Data (NLCD).

- Wildlife habitat (page A-23). This criterion also applies an extremely broad definition of wildlife habitat. Is open water really appropriately considered to be "wildlife habitat" given that it likely is an artificial reservoir in this region? The wildlife matrix measure also ignores effects from the landscape matrix. Adjacent land uses can reduce the quality and use of habitat by wildlife.

- Wildlife habitat (page A23) and wildlife habitat quality (page A-24). The differences between these criteria are not clear.

- Habitat fragmentation (page A-25). In extremely fragmented small-patch landscapes, average patch area and patch-area-ratio (PAR) can actually increase due to attrition of small patches. Additional description would be helpful. For example, is PAR calculated for all patches of any land use, or only for wildlife habitat? What will the scoring be for landscapes where wildlife habitat is the matrix and there are patches/perforations of disturbance? Is this measure appropriate for landscapes with long/linear riparian ecosystems that one would expect to see in xeric areas? Many of these important and critical ecosystems have high PAR. Also, given the pixel size used, are patches less than 10 ha in size omitted given the scale of the land cover data and analysis? This is important because the index might under represent the small patches and edge.

- Habitat fragmentation (page A-25). This criterion assigns a high value of vulnerability only to unfragmented lands. Habitat is simply lumped together as all undeveloped land types. Highly fragmented lands are also highly vulnerable too.

- Ecologically-significant stream segments (page A-29). It is not clear why stream segments defined in this criterion are "ecologically significant." The criterion is loosely and subjectively defined. A better description is needed.

- TEAP criteria (pages A-30 – A-34). Because these criteria are derived from EPA's Critical Ecosystem Assessment Model (CrEAM), the SAB assessment of CrEAM should be carefully reviewed (EPA Science Advisory Board, 2005) and appropriate recommendations should be applied to TEAP criteria. For example, the CrEAM review panel strongly cautioned against heavy reliance on "diversity." In GISST, diversity does not always indicate ecological vulnerability. In fact, some heterogeneous landscapes that are naturally fragmented may be more resistant/resilient to disturbance.

- TEAP criteria (pages A-30 – A-34). Using individual TEAP scores and the TEAP composite (which is the sum of the individual TEAP scores) is redundant.

- TEAP sustainability (page A-32). Sustainability is defined here as resistant to disturbance. If that is so, then why should the most sustainable sites be also considered the most vulnerable? Also, the value is based on the average of 30 m cells within $1km^2$ polygons. This will create artifacts from the resampling problem. This may be seen in other criteria as well but averaging rather than summing reduces the problem.

- Density of managed lands (page A-37). Density of managed lands includes

presence/absence only but does not reflect density or amount as suggested by the text. It is not clear how vulnerability would be measured.

- Colonias (page A-41). Colonias are included by counting numbers, but size is disregarded. In terms of understanding potential human health vulnerability, which is more meaningful – size or number? Is a single large colonia less vulnerable than several small ones? It seems that the actual population number is more appropriate.

- High school education and Educational achievement ranking (pages A-42 and A-43). These criteria are probably highly correlated, along with some other socioeconomic criteria.

- Environmental justice (pages A-44 – A-45). Why is a four square mile radius from a boundary of a facility applied for economic (environmental justice) measures? A statement of the rationale would be helpful.

- Minority, and other criteria where the state average is used, (page A-45). The use of state average data for the low end of the ranking score and two times the average as the high end does not allow a full examination of the factor. This approach truncates the distribution of data and in some cases cannot be used (e.g., if the state average is >50%). It would be more appropriate to use quantile scores.

- Socioeconomic measures that rely on percentages of populations showing a particular trait (pages A-42, A-44, A-45, A-46, A-48, A-50, A-51, A-53, A-56, A-57, A-58, A-59, A-60). In large metropolitan areas or heavily populated counties, there could be large numbers of affected people but this may represent a low percentage of the total population. Would absolute number be a better indicator in these cases? Examples of these criteria include: well water, telephone communication, linguistic isolation, and houses lacking plumbing.

- Age (page A-46). The criterion name does not reflect the purpose of the criterion.

- Age, Children, and Older population (pages A-46, A-48, and A-50). These criteria are closely related and probably highly correlated so why include all three?

- Phosphorus budget (page A-71). Phosphorus budget seems to be missing the first definition of how it is calculated (as included in nitrogen budget).

Appendix B. Specific Comments on the GISST User's Manual

Page 1, sentence 1: Improve the definition of the tool and be consistent in the use of this definition throughout the document. Be sure to make it clear how the GISST differs from traditional GIS use.

Page 2, line 4: Improve the description of the tool. For example EPA might want to say something like, "GISST is a prioritization tool. That is, given several options for a decision that affects the way land is used, the tool can be used to display a map with scores indicating potential ramifications of alternative decisions on the environment so the user can assess potential impacts of each decision and where vulnerabilities occur."

Page 2: Specify who can use GISST. Identify the kind of data, computer system, and computer expertise needed to use GISST. Specify the geographic region(s) where the tool can be used. Consider defining who the "users" are (e.g., EPA staff or stakeholders who examine the GISST output.)

Page 3, Line 5: Allow the user to establish the rules for combining the scores.

Page 4: While GISST may be transparent to the EPA staff that developed the tool, it is not clear whether the GISST is transparent to others. This point is made on page 70 where it is said, "for people not familiar with GISST or the output, this [initial spreadsheet] was not very user friendly."

Page 4: Make it clear if the user must specify the rules by which the criteria are combined.

Page 5: Each of the drawbacks should be discussed in more detail. The emphasis on the use of GISST as a screening tool is important and should be mentioned earlier in the document as well as in this section. Is the scale of resolution of the data a constraint?

Page 7, second sentence: Be consistent throughout the document in the way the tool is defined and described. The document might say, "The tool is an environmental assessment tool developed to provide more systematic approach *to using spatial data* to consider cumulative impacts in making environmentally sound decisions."

Page 13, second paragraph: The choice to use equal weighting of the criteria affects the outcome as well. A discussion of the implications of weighting is an important part of the document.

Page 14: The first paragraph needs expansion. Why are these properties key for assessment? It is not clear why the ability to manipulate spatial data is not mentioned as a key property.

Page 16: While the Panel was glad to see Table 1 included in the document, it was disappointing because the Table makes it clear that the GISST includes no information on ecological processes and disturbance regimes, and only limited information on biotic condition and chemical and physical features. Perhaps this long table should be an appendix and the key results should be reported in the text.

Page 20: Chapter 3 is much more than "Criteria Development."

Page 21: It is not clear how A_t is defined. Is it the actual area of the project (i.e. the road) or is there some buffer around the actual project?

Page 21: De-emphasize the primary algorithm as being a general case.

Page 21: It is confusing to use D_v to represent both individual criteria and the average of those criteria (and same holds for D_I).

Page 21, line 16: It is not apparent why "five options" are included in the comparative risk. What are these options? This paragraph is not clear.

Page 22: The report correctly points out the bias caused by using more than one criterion for a resource but does not suggest the basis for selecting the number of criteria to use, and how they are distributed among the resources. Perhaps the SAB Framework (EPA Science Advisory Board, 2002) can be useful in this decision by providing an approach for grouping sets of criteria.

Page 24: The concept of "flipping" criteria is important, but it is also important that criteria have a direction of impact with a lower number always implying less impact.

Page 25: It is surprising that the air quality section has few criteria. One would think that this section could be better populated with criteria.

Page 27: The application examples are important to include, but the peer review appendix is inadequate. It is not clear how the peer review was conducted. What was the basis for selecting reviewers? How were they engaged in the process?

Page 29 Last paragraph: It would be useful to know how the stakeholders were involved, not just who was involved.

Page 40, Table 2: Identify those data that were used in GISST in this application.

Page 48: The following statement on this page is unclear "The method described in the GISST User's Manual needed to be modified." The GISST User's Manual is describing a modification of the method in the User's Manual. Clarification is needed.

Page 48: The least cost path analysis sounds interesting, but the quantitative procedure is not fully described. How does this approach relate to the Transportation Problem of Mathematical Programming?

Page 76: The Help Sheet begins to provide information that is needed for a User's Manual, but it does not go far enough. More detailed information is needed.

The following changes are suggested for Appendices A and B of the GISST User's Manual:

- Clearer organization of the criteria and a preamble to explain this organization is needed.
- Hierarchical organization of the criteria is recommended.
- The peer review process is poorly described. The basis for selecting criteria is not provided. The quality assurance and control process is not documented. Additional information is needed in these areas.
- It would be better not to group "definitions, assumptions, limitation and uncertainties" but instead to provide each separately. It would be useful to provide a single list of the criteria in Appendices A and B.
- A more complete description is needed for each effect and for the criteria used to provide data on that effect.
- Although the GIS data sets utilized for the GISST are existing data sets, better documentation is needed for the source data, for the benefit of future users and to demonstrate understanding of the data set derivation.
    - URLs should be provided for every data set. The URLs should point to the original data sources and/or metadata that have been written for those data sources.
    - For each data layer there should be a description of the data by the source agency (e.g., the spatial aggregation of census data used in the GISST was by census block).
    - If EPA Region 6 summarized the data by a geographic unit (e.g., HUC) after the data were obtained from the source agency, the procedure used to summarize the data should be documented.
    - The time frame of the data (e.g., NPDES permits issued in 2004) should be provided.
- A description of scaling of data should be provided.