DRAFT – DO NOT CITE R QUOTE – This draft SAB letter is a work in progress. It has been prepared for final review and approval of the Chartered SAB. This draft letter does not represent EPA policy. DECEMBER 11, 2006

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

OFFICE OF THE
ADMINISTRATOR
SCIENCE ADVISORY BOARD

_____ , 2006

The Honorable Stephen L. Johnson
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington D.C. 20406

The Honorable George Gray
Assistant Administrator ORD
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington D.C. 20406

In 2006 the Science Advisory Board reviewed the EPA Region 6 Compliance Assurance and Enforcement Division's Geographic Information System Screening Tool (GISST). GISST is a tool developed by Region 6 for rapid preliminary environmental assessment and reconnaissance in support of environmental impact statement evaluations. The SAB report on that review was forwarded to you on September 28, 2006 (*SAB Review of the EPA Region 6 Geographic Information System Screening Tool*; EPA-SAB-06-011).

The SAB has also received briefings on Region 6 and Region 4 activities in response to hurricane Katrina including application of GIS tools to target sampling efforts and assist in clean-up operations. The SAB has also reviewed or been briefed on a number of other GIS based tools developed by other EPA regions, laboratories and headquarter's offices. Examples are the Region 4 Southeastern Ecological Framework, the Region 5 Critical Ecosystem Assessment Model, the Office of Water's Index of Watershed Indicators, and the ORD Regional Vulnerability Assessment Methods.

Discussions among SAB members during its review of the draft SAB Report on the Region 6 GIS based screening tool (GISST) catalyzed a broader discussion of a systemic problem that the SAB has observed with other EPA developed GIS based decision support tools. That is, a number these tools do not adequately utilize modern decision analytic techniques in their development of numerical scoring. Single vulnerability or impact scores if improperly derived can mask important differences in individual data layers critical to the evaluation of environmental impacts. It appears that developers of these GIS based tools ignore or are unaware of the large literature on multi-attribute decision making.

It is too expensive to send a team of experts to assess every plot of land for which a change in land-use might be considered. Thus, models which utilize broadly available remote-sensing data, and perhaps other comprehensive and spatially specific information, are desired in an attempt to reproduce the types of ratings that might be assigned to particular plots of land by experts if they had the opportunity to study the plot in person. From a statistical perspective, the way to construct such a model is to select a sufficiently large random sample of plots of land and to actually send teams of experts to each plot to make a quantitative evaluation (probably along a number of several distinct dimensions). These judgments can then be treated as dependent variables in statistical models that seek to explain systematic variation in these actual (and preferred) on-the-ground expert judgments as a function of whatever information can be assembled from remote-sensing or analogous data. Once the model has been statistically calibrated--and if it can be shown to reliably explain a substantial amount of the observed variation in expert judgment across different types of plots--then there is some basis for arguing that one can rely upon models and remotely sensed data to replace the on-site judgment of a team of experts.

Currently, the models which are in development do not rely on representative samples and statistical inference. There is no opportunity to allow the data to dictate the most appropriate functional forms for these models (i.e. the formulas which should be used to combine the remotely sensed data to produce an indicator of the status of a given plot of land). Instead, these formulas typically involve arbitrary weights on different types of remotely sensed data, based upon intuition, and in some cases, spurious normalizations—a strategy that could be described as "guessing at regression coefficients in the absence of a dependent variable." The weights being used can be described, at best, as testable hypotheses about the relationships between remotely sensed data and the status of any particular plot of land. But these hypotheses are merely that. They have been neither confirmed nor refuted with real data, although they could be.

Considering the growing use of GIS based assessment tools in EPA, the diffuse origins of such tools within EPA and their potential value to assist EPA decision making, the SAB strongly encourages the Agency to utilize the guidance of its Council on Regulatory Environmental Modeling (CREM).. It is very apparent that technical assistance needs to be provided to developers of GIS based decision support systems on regression modeling and other analytic techniques.

Sincerely

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